47I'

MODERN PHYSICS

Digitized by the Internet Archive
in 2023

https://archive.org/details/modernphysicsgen0000theo

MODERN PHYSICS

A GENERAL SURVEY OF ITS PRINCIPLES

BY

THEODOR WULF, S.J.

IGNATIUS KOLLEG, VALKENBURG, HOLLAND

TRANSLATED FROM THE SECOND GERMAN EDITION BY

C. J. SMITH, Ph.D., M.Sc., A.R.C.S.

LECTURER IN PHYSICS, ROYAL HOLLOWAY COLLEGE, ENGLEFIELD GREEN, SURREY

BRISTOL PUBLIC LIBRARIES
Reference Library

BRISTOL CITY LIBRARIES
WITHDRAWN AND OFFERED FOR SALE
SOLD AS SEEN

WITH 202 DIAGRAMS

METHUEN & CO. LTD.
36 ESSEX STREET W.C.
LONDON

0
554

Originally Published in German under the title
"Lehrbuch der Physik"
This Translation First Published in 1930

PRINTED IN GREAT BRITAIN

FROM THE PREFACE TO THE FIRST GERMAN EDITION

A TREATISE which had the title 'Elementares Lehrbuch der Physik' was published more than thirty years ago by my deceased teacher and predecessor, the **Rev. Father P. Ludwig Dressel.** In his time considerable appreciation was bestowed upon the several editions of the book which had a wide circulation. For several years the work has been out of print, but the publishers now desire to republish it.

In the time which has elapsed since Dressel's book was written profound changes have taken place in the world of natural science. In a book dealing with the present position of physics these marked changes must make themselves felt, although textbooks can only follow the march of events at a certain distance behind.

The alterations consist partly in the addition to the traditional subjects of several new and important chapters on such subjects as radioactivity, X-rays, spectrum analysis, and the theory of relativity. These new sections could have been appended to the usual matter without any very great difficulty—in fact, this was done by **P. Paffrath**, who revised the last edition of the book very carefully. But the fields of research which have recently been opened up have given such an impetus to the train of physical thought, that its centre of gravity has been considerably shifted. For whereas the law of the conservation of energy formed the basis for all subjects considered in the book by **Dressel**, the structure of the atom has long since become the goal towards which the attention of research workers has been especially directed. It therefore appeared desirable that this altered viewpoint should find expression throughout the whole of the new book. A third change which has since been made must also make its influence felt throughout. Until last century the main task in the study of physics, even if it were not its sole aim, was to describe natural phenomena, and in general to summarize them in formal laws. To-day it has set itself another task ; it seeks to explain all phenomena in terms of the small atomic particles which constitute the material world. Its endeavour has been to describe the events completely, so that it might go beyond a purely phenomenalistic conception of things ; for to describe a whole series of facts without making some attempt at a definite hypothesis concerning the inner mechanism of the material atoms is not very satisfactory. We may illustrate these remarks by referring to the behaviour of gases, and especially of the small deviations which they exhibit from the laws of perfect gases : also osmotic pressure and the exceptional behaviour of electrolytes. Again, we may cite the confusing phenomena which arise from the passage of electricity through rarefied gases, the processes which take place in radioactive substances, and many others. Now, whereas experiment only yielded a phenomenalistic description and not a satisfactory and comprehensive representation of things, research made great and rapid strides when the events in the material world were viewed from an atomistic standpoint. At the same time an explanation founded upon atomic ideas was much more satisfactory, because it gave an insight into the relationships which exist in matter. With the possibility of arriving at an understanding of natural processes, the desire for it arose with the irresistible impulse of natural instinct. If the atomic structure of matter was at first the means of giving a systematic view of physics, it soon became the subject for research itself. The very successful development of physics, which we have experienced since the beginning of the present century, consists mainly in a successful explanation of atomic phenomena. **The derivation of phenomena on the basis of the internal structure of the smallest constituent particles of matter is considered to be an essential and final goal of physics.** The attainment of this object is the aim to which all physicists from all parts of the globe have directed their attention along all possible channels, and with all the knowledge and powers which research has hitherto placed at their disposal.

This profound change in the aim of physics must necessarily lead to a far-reaching alteration of the whole exposition, if this is to reflect how research is progressing. I have therefore decided to write quite a new book, and this I now present to my readers. But my aim is still the same as that which was in the mind of my esteemed teacher, namely, **to give a coordinated account of the fundamental results of physics.**

Such coordinated accounts are needed more than ever at the present time. The great developments which have taken place in physics during recent years have given rise to special researches which are so extensive that it is not possible, even for an expert, to have a first-hand acquaintance with all this work, whereas the non-technical person, who must utilize and consider the results of physics in his daily work, is entirely dependent upon such a summary if he is to obtain the necessary ideas.

It is essentially for those readers, who are not technical experts, that the book has been written. The number of these who, without wishing to become physicists, desire to obtain an insight into the realm of physics, is to-day very great, and it must become more numerous the further physical research is directed towards the intrinsic happenings of the material world whence all phenomena, including those discussed in other branches of science, take their origin. Amongst my readers I also count upon many who studied physics some years ago, and who now wish to supplement their knowledge without having to read the original literature ; I have remembered also engineers, teachers in secondary schools, and that multitude of educated people in general who wish to become acquainted with the results of physics. A mere enumeration of facts or a repetition of prevalent conceptions is of little use to such persons. If a reading of the book is to prove at all profitable to them, it must afford them some idea concerning the interrelationships of the various phenomena ; they must know, and in a certain measure learn how to appreciate the facts from which our modern ideas have sprung. I have endeavoured to develop these connexions, in so far as the present position of research and my own powers have permitted.

Another class of readers must also be mentioned ; it includes the representatives of philosophy. For much too long a time physics and philosophy have been at enmity with each other ; philosophy has frequently concerned itself only a little with the results of physics, and physics, on account of this indifference towards physical results, has viewed philosophy with a certain amount of scorn. As matters stood, it could hardly have been otherwise. For every branch of philosophy saw its main task in trying to discover connexions which would lead to an understanding and comprehension of the processes of Nature. As long therefore as physics refused to discuss these relationships, these two subjects could scarcely have much in common. In the last few years, however, matters have changed considerably. For it is now just that original problem of philosophy concerning the structure of matter to which physics, with all its great powers, has directed its attention. Henceforward the achievements of physics can no longer be ignored by philosophers. In fact, all intelligent advocates of philosophy, who wish to build upon facts and not construct air-ships from the boundless realms of fantasy, have since endeavoured to become acquainted with the results of physics. To such people this book ought to be helpful ; it might lighten their task in their choice of what is important and in their understanding of what is difficult, since it is based upon the lectures which I have given to students of philosophy for more than twenty years. In this book those results of physics will be given which might serve to explain natural phenomena from the viewpoint of modern research, and which therefore must be considered by philosophy, if it makes claim to be a science founded upon experience.

This aim has been important both for the choice of the subject matter and for the manner in which it is presented. With regard to the contents, all that might have served more in technical applications than in a comprehension of fundamental ideas has been omitted.

By this arrangement the really fundamental idea, viz., the structure of matter, to which everything else is quite subordinate, stands out clearly and distinctly. Everything is directed to this goal, and although a few sections of a somewhat preparatory nature are necessary, they are treated as briefly as possible. But the deeper a question penetrates into the structure of matter, the more it tends to transform our existing conceptions, and hence it will be more thoroughly discussed and the proofs more completely given, so that the reader can form for himself an opinion concerning the importance of the fundamental facts. That, at any rate, was my intention : how far I have succeeded in doing so, must be left to the judgment of those readers who are familiar with this subject.

The manner in which the subject is presented was determined by the purpose of the book. It enables the non-physicist to understand the fundamental ideas of physics, and I realize quite well that a clear line of demarcation must be drawn between things proved, and suggested explanations of the facts. Technical expressions, which are not easily understood, have been avoided wherever possible, and an explanation capable of being comprehended by any intelligent reader has always been sought. Long mathematical proofs have not been given, but this does not mean that mathematics has been rejected on principle.

Consideration of the purpose of the book finally requires in certain cases a departure from the usual method of presentation so that one does not burden the reader with doubt and difficulties in an introductory course. For readers who desire a correct estimate of what has been attained it is imperative to consider certain differences and difficulties,—if for nothing else on account of the truth. For these difficulties are in some places of quite a serious nature and one must not lose sight of the fact that the views now holding may again experience a fundamental change. In the application of the results to a neighbouring field it is often important to know how exactly the validity of a principle or law holds, and how far we can rely upon the evidence of the facts available. But in order to avoid dissatisfaction in the reader the confirmations must be presented in as true and compatible a manner as possible. In estimating the pros and cons I have always done my best, and in so far that I have regarded certainties as certainties, suppositions as suppositions, and unsolved problems as unsolved, I believe that the cause of truth has best been served.

T. WULF

July, 1926

TRANSLATOR'S NOTE

THIS translation has been made in accordance with the second German edition (1929), and except in one or two places the German text has been preserved throughout.

My very best thanks are due to Professor C. H. Desch, F.R.S., for his help, and to Dr. R. W. Lawson for his numerous suggestions and advice. Also to Messrs. Edward Arnold & Co. for kind permission to reproduce Fig. 163, from my textbook *The Principles of Physics for Students of Pharmacy*, and to Messrs. E. Leybold's Nachfolger A.G., Cologne, for the loan of blocks for Figs. 23 and 151.

C. J. SMITH

SHEFFIELD
August, 1929

CONTENTS

PAGE

Introduction 1

PART I

THE MATERIAL WORLD

CHAPTER I. Fundamental Ideas concerning Motion.

1. Measurement and Units 7
2. Motion 12
3. Velocity 17
4. Acceleration 21

CHAPTER II. The Motion of Bodies.

1. The Teachings of the Ancients concerning Motion 23
2. The Laws of Motion according to Galileo and Newton 24
3. Momentum and Impulse 33
4. Action and Reaction 34
5. Work and Energy 37
6. Motion in a Circle 41
7. Rectilinear Harmonic Motion 44

CHAPTER III. Gravity and Gravitational Fields of Force.

1. Gravity 48
2. Gravitational Fields of Force in General 50
3. Potential 56
4. The Nature of Gravitational Attraction 59
5. Inertia and Gravitation 62

CHAPTER IV. Elasticity 64

CHAPTER V. Impact 69

CHAPTER VI. Waves.

1. Fundamental Ideas 73
2. The Transmission of Waves 79
3. The Disturbed Propagation of Waves 83

CHAPTER VII. Acoustics.

1. Sound and Music 88
2. The Propagation of Sound 93
3. Sources of Sound 98

PART II

THE ATOMIC STRUCTURE OF MATTER

CHAPTER I. The Classification of the Elements 105

CHAPTER II. The Gaseous State of Aggregation.

1. Fundamental Facts 113
2. The Kinetic Theory of Gases 116

CHAPTER III. The Fluid State.

1. Fundamental Properties of Liquids 128
2. Molecular Movements in Liquids 131

CHAPTER IV. The Solid State 136

CHAPTER V. Temperature 141

PAGE

CHAPTER VI. **Specific Heat.**

1. The Idea, Unit, and Determination of Specific Heats . . 150
2. The Specific Heats of Gases 152
3. Atomic Heats 157
4. The Quantum Theory and Atomic Heats 159

CHAPTER VII. **Heat and the Various States of Aggregation.**

1. Preliminary Remarks on the Change of State . . . 162
2. Fusion 164
3. Evaporation 166
4. Boiling 171
5. Critical Temperature and Pressure 172

CHAPTER VIII. **The Laws of Thermodynamics.**

1. The Conservation of Energy and the First Law of Thermodynamics 174
2. The Second Law of Thermodynamics—Entropy . . . 185
3. The Third Law of Thermodynamics 197

CHAPTER IX. **The Atom.**

1. The Existence of Atoms 201
2. The Determination of Atomic Constants 204
3. The Distribution of Molecular Velocities 212
4. Isotopes 215

PART III

THE STRUCTURE OF THE ATOM

Introduction 221

CHAPTER I. **Magnetism.**

1. The Fundamental Laws of Magnetism 222
2. Theories of Magnetism 229

CHAPTER II. **Electrostatics.**

1. Fundamental Phenomena 233
2. Electric Potential 235
3. Electrification by Influence 238
4. The Effect of the Medium 241

CHAPTER III. **The Electric Current.**

1. The Laws of Current Electricity 246
2. The Theory of Metallic Conduction 251
3. Thermoelectricity 253

CHAPTER IV. **Electromagnetics.**

1. Magnetic Effects due to a Current 256
2. The Systems of Electrical Units 262
3. Electromagnetic Induction 264

CHAPTER V. **Electrolysis.**

1. Fundamental Facts 270
2. Theories of Electrolysis 278
3. The Ions of Electrolytes 282
4. Galvanic Cells 284

CHAPTER VI. **Ions and Electrons.**

1. Cathode Rays 287
2. The Determination of ε 292
3. The Electron 294
4. Positive Rays 299
5. Gaseous Ions 301
6. The Passage of Material Rays through Solid Substances . . 306

PAGE

CHAPTER VII. **Radioactivity.**

1. THE DISCOVERY OF RADIUM 308
2. THE RAYS FROM RADIOACTIVE MATERIALS 309
3. RADIOACTIVE SUBSTANCES 312
4. DEDUCTIONS FROM RADIOACTIVE PHENOMENA 317

CHAPTER VIII. **Electromagnetic Waves.**

1. ELECTROMAGNETIC WAVES 328
2. HEAT RAYS 331
3. PLANCK'S QUANTUM THEORY 336
4. LIGHT RAYS 340
5. X-RAYS 358

CHAPTER IX. **The Structure of the Atom.**

1. ELECTRONS 368
2. THE NUCLEI OF ATOMS 376
3. ATOMS AND MOLECULES 380

PART IV

THE PHYSICS OF THE ÆTHER

INTRODUCTION 383

CHAPTER I. **The Undisturbed Propagation of Waves in the Æther.**

1. THE RECTILINEAR PROPAGATION OF LIGHT 384
2. THE VELOCITY OF LIGHT 384
3. THE NATURE OF THE TRANSMISSION OF LIGHT 387
4. THE MEDIUM IN WHICH LIGHT TRAVELS 389

CHAPTER II. **The Disturbed Propagation of Light.**

1. THE REFLEXION OF LIGHT 392
2. THE REFRACTION OF LIGHT 398
3. THE THEORY OF DISPERSION 403
4. OPTICAL INSTRUMENTS AS AN AID TO VISION 405

CHAPTER III. **Interference and Diffraction.**

1. INTERFERENCE PHENOMENA 408
2. THE DIFFRACTION OF LIGHT 415

CHAPTER IV. **The Polarization of Light.**

1. POLARIZATION BY REFLEXION. 425
2. POLARIZATION BY REFRACTION 426
3. DOUBLE REFRACTION 427
4. POLARIZATION BY DOUBLE REFRACTION 429

CHAPTER V. **Light and Colour.**

1. NATURAL LIGHT 432
2. COLOUR 434

CHAPTER VI. **The Æther.**

1. THE EXISTENCE OF THE ÆTHER 437
2. THE PROPERTIES OF THE ÆTHER 439
3. THE MOTION OF THE ÆTHER 441
4. EINSTEIN'S THEORY OF RELATIVITY AND THE ÆTHER . . . 447

CHAPTER VII. **The Aims of Modern Physics.**

1. LONG ESTABLISHED FACTS 454
2. NEW FACTS 458
3. MODERN AIMS 461

INDEX TO SUBJECTS 465
INDEX TO AUTHORS 467

MODERN PHYSICS

INTRODUCTION

THE struggle of mankind with the secrets of Nature is certainly as old as the human race itself; in fact Nature was the first thing against which man had to battle for his existence. First with her beauty and grandeur, then with her abundance and liberality, and finally with her formidableness and majestic awe; these were the means whereby Nature came into touch with the human race.

It was impossible for man to be with Nature and not pay heed to her call, for he was dependent upon her, and without her he could not eke out even a bare existence. He needed food, clothing, and a home, and only Nature could provide him with these necessities. At times her gifts were offered quite readily and willingly, but, generally speaking, hard work had to be done to render serviceable the things she had to offer. Sometimes it happened that Nature, in angry mood, would destroy in an instant everything that he had procured for himself after years of toil; a flood engulfed it, a fire devoured it, a storm razed it, or an earthquake scattered it in confusion, instantly converting huts and houses into a heap of ruins.

Afterwards, ways and means were sought whereby such calamities could be avoided, for even when the danger was over and life's pressing needs satisfied, there still remained a vivid recollection of what had happened and an innate desire for knowledge and understanding caused man to ponder deeply over the hidden secrets of Nature.

The sequence of events was noticed and very soon afterwards a regularity in this succession was bound to be discerned: the uninterrupted change from day to night in conjunction with the daily transit of the sun across the sky; the differing lengths of the day at the various seasons of the year and the corresponding changes in plant life and the sun's altitude; the impetus of a projected stone or spear in relation to its mass and velocity.

Man's desire for understanding, however, was only satisfied when he was able, or at least thought that he was able, to ear-mark some event as the cause of another—the exertion in his arm was the cause, the flight of the spear was the effect.

He was further guided in his inquiries by his desire for general laws. Quite silently and naturally he was convinced that under similar conditions similar events would happen. He believed that, what he had come to know concerning one thing, he would find again under similar circumstances with other things. He realized that a spear would travel the same path every time he hurled it with the same velocity in the same direction, for otherwise there would have been no sense in practising the art of throwing a spear.

These two principles, first the regularities which are found in all natural phenomena, and secondly the effective cause to which every happening must be attributed, have been like two natural mental stimuli to every man, in whom, consciously or unconsciously, they exist. They are the two pivots around which his thoughts must centre.

A third factor, too, was soon added to the need of man and his thirst for knowledge, and that was his desire for gain—the 'auri sacra fames'. Sagacious minds had soon become sure that knowledge meant might and wealth. At first, knowledge gave man power over the forces of Nature. It was discovered how the unbridled powers could be subdued and made submissive so that these, instead of causing ruin, would further man's aims. He discovered how to make use of the wind, water, and fire, in his machines, how these could quicken the dispatch of news, and how to obtain a service from them, which repaid him a thousandfold. Moreover, the investigation concerning the secrets of Nature became an ever-flowing source of unprecedented richness. The further knowledge progressed the more extensively could industry expand, until, to-day, millions of workers earn their daily bread by it, and at the same time it offers to an ever-increasing army of research workers means which, in their turn, again benefit industry and augment the resources for further progress.

We shall only concern ourselves with the results of this research in so far as they satisfy our desire for knowledge. We see the outside world running its course, we are confronted with a thousand problems and suspect a kingdom where law and order reign. We want to make this world a mental possession of ours ; we want to procure an inner world which exists simultaneously with the external one, and which shall be a true picture of the world around us. But this world inside us will not be sufficient if we can only observe the striking of the hour from afar, for we wish to see the inner workings of our 'world clock', and understand the reason for the silent, but ever forward, motion of its fingers. It is that for which we long, and that alone can satisfy our desire for knowledge. Such a work, however, is not to be constructed from superficial observations made by some amateur, but must be found out by serious intellectual endeavours. Just as an artist glances repeatedly upon his subject, bettering and improving his picture accordingly, so must we also compare the picture of Nature which we have conceived with the actual thing, and, according to the comparison, we must improve and complete the picture so that it shall resemble the subject more closely.

Natural Science and its Progress through the Ages. Concerning the activity of research itself, two periods must be distinguished ; they will be called the ancient and modern periods. The modern period begins with the close of the Middle Ages. The fact that there are two periods does not imply that the aim of research has been different in the two periods. At all times the objects of research have been the same, viz., to obtain the most complete and profound knowledge of Nature, *i.e.* the whole material world. But the ways which were chosen, at different times, to reach this goal differ considerably ; and yet, in principle, the ways were not different. It had always been realized that the commencement of investigations concerning Nature could only be made from the observation of things, and that one must start with the observation of individual phenomena in order to arrive at a general law. But the importance of this conception was assessed at very different values in the two periods.

In the ancient period it was believed that a few superficial observations, mostly made by chance upon events which took place on their own accord, sufficed as a basis for a discussion concerning the nature of things.

For the purposes of this discussion one was very often content just to relate the observations which had been made by others. Without paying much heed to the veracity of the reports the discussion concerning the nature of the phenomena was soon commenced, so that the conclusions which had been made without sufficient support had later to be discarded.

In the **modern period**, however, the procedure was different.

1. The question as to the nature of things was, at first, held in the background, and pains were taken in order to obtain a description of all individual processes which should be as complete and universal as possible, so that, later on, the general laws which govern the phenomena, and the nature of the participating bodies, could be discussed.

2. The desire to obtain exact information concerning the why and wherefore of a process is so great that one is no longer content to observe chance events, but natural processes are intentionally produced so that one may observe them. These latter are called experiments, whereas, by observations, we understand the investigation of such processes as occur spontaneously, *i.e.* without human intervention, and due solely to influences available in Nature.

The important rôle played by experiments in modern physics is so great that we must pause a moment. The chief difference between experiment and observation lies in the fact that in an experiment the observer himself determines all circumstances of place and time, selects all the influences which impede or promote the process, and then directs his whole attention carefully to the specifically doubtful point, whereby his observational powers are augmented where necessary by suitable contrivances such as the telescope, the microscope, photography, automatic registration, the cinematograph, etc.

A very effective way of establishing the validity of a conclusion consists in the frequent repetition of an experiment, so that a certain factor, the influence of which is under investigation, is varied within wide limits, from the smallest to the largest value practicable. The results will show whether or not the factor in question influences the phenomena under examination.

3. All this work assumes a greater significance since, in contrast to the earlier qualitative observations, the process is now investigated quantitatively, *i.e.* the influence of all factors is ascertained numerically as far as possible. In this way laws have been obtained from which numerical predictions concerning natural phenomena can be made.

Such definite laws are of great help when it is necessary to deduce, from the observations, conclusions relating to the causes of the phenomena. Now a conclusion made from one phenomenon is, in general, not unique, for generally there may be several different causes which may be responsible for the phenomenon under consideration. The different characteristics of the process, which are revealed by experiment, represent just as many conditions as must be satisfied by the cause. The more numerous conditions, each of which must be present for the process to happen these requirements, the less numerous will be the causes which may be responsible for the effect in question.

The success achieved has shown that the modern methods for the prosecution of research are effective to an extent which was never contemplated before, and, by utilizing these means for research, physics has grown very rapidly since the end of the Middle Ages, and its rate of progress is still increasing.

The Scope and Aim of Physics. An immediate consequence of this progress was the necessity for an extensive subdivision of the subject. Whilst some workers have endeavoured to extend our knowledge concerning the forces and the properties of matter in general, others have concentrated their attention upon special branches of the subject, *e.g.* plants, animals, rocks, the stars and the earth. Whilst they have been extending their knowledge, and perfecting their methods of research, they have discovered new means of research which have often been particularly suited for that special line of inquiry. Thus the progress of their own research has been accelerated, but others have found it difficult to become acquainted with the progress in these other directions, or to acquire the skill necessary to use the more perfect instruments. Natural science, which, until then, had represented the vast domain of all nature, now divided itself into a large number of subsidiary subjects. We may not mention astronomy, botany, zoology, mineralogy and geology. Only two branches of natural science, viz., physics and chemistry, concern themselves with the universe as a whole. Both these sciences examine the same universe from a different aspect. It has been proved that most of the bodies which occur in Nature can be decomposed into a number of substances or elements, essentially different from one another, and not capable of being further subdivided to form new elements. The number of these substances is extraordinarily small when it is contrasted with the manifoldness of the material world. At the present time about ninety are known. The decomposition of bodies into their ultimate constituents (analysis) and the formation of compounds from these elements (synthesis) are the object of chemistry. If the products obtained from synthesis or analysis are called new substances, we may define chemistry as that branch of knowledge which discusses those phenomena connected with the changes of substance which occur amongst material bodies.

On the other hand physics discusses the large number of material processes which occur in Nature, but which are not associated with such profound changes in the nature of the participating bodies, and one might well imagine that such a study would not permit us to penetrate very deeply the inner secrets of Nature. Actually, however, it is from just such a study that, in our own time, a whole series of discoveries have been made which have furthered our knowledge concerning the material world to an extent almost without precedent. The progress which has been made has attracted to this field of knowledge, in all countries, a whole series of prominent research workers who, whilst being active rivals, have helped one another to become more efficient workers. Even at the present time we are in the midst of a classical era where physical research is concerned; a time full of success, and yet one in which many quite new and more fundamental riddles still surround us. He who wishes to acquire a well-founded conception concerning the natural world cannot overlook these results, for otherwise he would be like the traveller who wishes to obtain an idea concerning the life and activities of a town by sleeping there for one night only.

Scope of the Book. In the present book those results of physical research which are of importance for an understanding of Nature have been compiled in a short and, in so far as the author was capable, a popular presentation.

The whole can be arranged in four sections. The first section concerns bodies as they appear to us and as we perceive them directly by means of our eyes. By further research we have learnt that these bodies are built out of very small particles which we call atoms. Even though these atoms cannot be seen directly with the eye, yet we still have quite certain proofs of their existence and of many of their properties. The whole of our knowledge concerning the atomic structure of matter has been discussed in the second section. The work of the last few decades has allowed us to penetrate still further into the inner structure of matter. The structure of the atom itself is examined in the third section. Finally there exists a whole series of phenomena, which have only been explained on the assumption that, in addition to ponderable matters, there is still something else which fills all interstellar space, a medium which is called the **æther**. The phenomena which take place in this medium are discussed in the last section of the book.

PART I

THE MATERIAL WORLD

CHAPTER I

FUNDAMENTAL IDEAS CONCERNING MOTION

1. Measurement and Units

FROM what has already been stated it follows that measurement must play an important part in all physical research. Not only is it essential for us in order to be able to describe natural events accurately, but still more so in order to answer the question as to how far our conceptions approach the reality.

The Measurement of Length. Let us now consider, as our first theme, the measurement of a length, for this requires no knowledge of physics, and historically it was the first process to be thus discussed. Let us assume that a hut is to be built, and that the necessary wood and beams for it are sought in the forest, where they are cut into shape. According to the purpose of the hut its length and breadth will be determined and indicated by means of stakes in the ground. Afterwards, a piece of string is obtained and, assisted by a colleague holding one end of it by one stake, the string is made taut and cut off at a point where it touches the second stake. In a similar way pieces of string representing the breadth and height of the hut are made. It is easily seen that measurements made in such a manner are really free from all objection if such a material is selected so that its length is always the same. **For the final purpose of all measurement is that afterwards, in other places and at other times, a length can be specified which will have the desired magnitude.** But this particular piece of string cannot state more than that. It takes no account of the breadth or height of the hut, nor of the colour or strength of the material. It is also quite immaterial what is the colour or the thickness of the string. It can only represent one thing, and that is the length of the hut. The string is only accurately adjusted to the length ; it takes no account of anything else whatsoever.

But it is easily understood how extraordinarily difficult and troublesome this process will be, when many such measurements are to be made and recorded ; still more so when others have to be informed about them. In addition one imagines how difficult it would be to estimate long distances by such a process.

An enormous advance in the science of metrology was made when it was recognized that a long length could be defined by means of a unit and a number, which indicated how many times the unit must be placed end to end in a straight line in order to attain the desired length. The name of the man who was the first to make such a measurement has not been handed down to us. We can imagine that, at first, a hut was to be built in which a definite number of strides could be made. The length of one stride was thereby chosen as the unit of length. Soon this rather indefinite length of a stride was replaced by that of the foot, which was better defined, and then, at a later epoch, a rod was chosen as the standard foot which was to be a basis for all measurements of length. Now at first, everything that had to be measured was connected in some way with human beings—the house to live in, the cloak to wear, the bed for sleep, the tools for work, the fields to cultivate—it is therefore not surprising that the first units were related to some feature of the human body—the foot, the ell, the fathom, the acre, and an hour's journey. The fathom is equal to the distance which can be encompassed when the arms are outstretched, whilst an acre was the area of land which could be ploughed in a morning. Each individual industry chose these primary units independently, and without concerning itself with another. Thus, until 1875, there were in existence the Prussian, Rhineland, English, and the French foot. Even in the same country different units were used for different purposes. The mariner used knots* and fathoms, the surveyor rods, the weaver ells, whilst the carpenter worked with feet and inches.

This multitude of independent units was bound to cause inconveniences, especially when the introduction of the railway, steam ships, telegraphy and telephony, caused the people of different countries to exchange their material as well as their intellectual goods with one another. Finally, in the year 1875, most of the civilized world chose the metre as the unit of length. The metre was to be the ten-millionth part of the earth's quadrant, *i.e.* the shortest distance from the North Pole to the Equator. A costly bar of platinum iridium, an alloy which is sufficiently hard and which possesses the power of resisting atmospheric corrosion, was manufactured in the form shown in Fig. 1. The distance between two fine marks engraved upon the central portion of the bar was declared to be the length of the standard metre. From this bar a series of similar shaped substandards were then prepared and given to different countries.

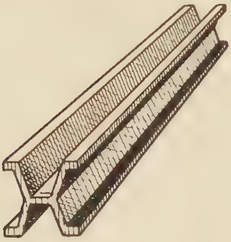


Fig. 1.

Now it is not the object of this book to discuss the science of measurement more closely. Founded on certain principles a length can always be defined by means of a number, which shows how frequently the metre rods must be placed end to end in order to attain the length of the measured object. We say, for example, 'This table is two metres long.'

The hundredth part of a metre (*m.*), the centimetre (*cm.*), has become the usual unit for scientific purposes. Still smaller dimensions are measured in millimetres (*mm.*), the thousandth part of which is called 1μ, whilst the millionth part of a millimetre is called $1\mu\mu$. Longer distances are expressed in kilometres (*km.*) and finally in light seconds and light years. Since light travels 300,000 *km.* in a second, a light second represents this distance. The magnitude of a light wave is almost invariably given in Ångström-units, where $1 \text{ Å} = 10^{-8}$ cm. $= 10^{-7}$ mm. The still shorter waves which we meet in X-rays are measured in X units, where $1 \text{ X} = 10^{-10}$ mm.

The Reliability of the Measurement of a Length. If, in carrying out a measurement, all possible care is taken to place the metre rods end to end in a straight line, without any gaps, are we absolutely sure of the validity of our result ? Now the measurement of a length is carried out as follows : the metre rod is placed alongside the object, and then moved on again. We must therefore inquire as to whether the metre rod changes when it is thus shifted. And it is only in so far as we are sure that the rod remains unaltered after this shifting, that we can be sure of the validity of our measurement. We must also ask whether such a rod can change when it is moved in space. This may happen if, at different localities, different physical influences act upon it. We may suppose, for example, that the temperature varies from place to place. At higher temperatures bodies ordinarily have a greater length. But we are able to recognize such physical changes and eliminate them.

---

* A 'knot' is a speed of one sea-mile (6,080 ft.) per hour. A sea-mile is one minute of arc along any great circle of the earth. The British Admiralty fixed it at the above value. [Trans.]

There remains yet another question. Can the shifting in space itself affect the metre rod ? Now if all bodies at different points in space are elongated or contracted proportionately, we have no means of ascertaining such a change. In fact, we have no reason to suppose that such an alteration can take place. In addition, we have never observed that space, as such, produces any effect at all on bodies. Expressed more exactly, we may say, that where an effect has been observed in Nature, we have always succeeded in locating the action of some real body as the effective cause of this process. But when we ask whether we are certain that all bodies do not change in a similar manner, when at different points in space, we have to reply that we are not certain about it.

Another question is always intimately associated with this one : Can we discover, by measurement, the **absolute** length of a body, or is it only one which is **relative** to the measuring rod ? It is easy to see that every measurement which has been made on the lines above consists essentially of a comparison with a standard rod, and is therefore relative. But this relative measurement would become an absolute one immediately, if we were sure of the absolute length of the measuring rod. **Boscović** in 1758 propounded this question to himself : ‘ What would happen if, one morning on awaking, we found that all bodies including our own had elongated in a definite ratio ? ’ He concluded that it would not be noticed. We therefore have no certainty concerning the absolute length, either of objects, or the measuring rod itself, which, after all, is only an object. What we can specify as the result of our measurement is only the length compared with that of the measuring rod. By this we do not mean that the absolute length of bodies will change from time to time. We can neither conceive a satisfactory basis for such a change, still less since all bodies would in general be involved at the same time and in the same way, nor have we any ground for surmising that it might have happened in the past. The next question is only a theoretical one. Do we ever find in Nature itself an actual state of affairs from which we can conclude with certainty that such a change has not taken place ? This question has to be answered in the negative.

The Measurement of Time. Every man observes in himself not only a large number of sensations, humours, sentiments, wishes and desires, but he has convinced himself that the different personal experiences do not exist simultaneously, but that they follow one another, to some extent limiting and causing one another. Whilst I am working at my writing desk I am surprised by the arrival of a friend ; then follow the *tête-à-tête*, the exchange of experiences, the lunch together, the departure of my friend, and finally solitude, with the impressions which have been received.

In return we have a certain feeling that between the friend's arrival and departure several things have happened, on one occasion more, on another less. One time there is only a short greeting, another time he stays several nights. And thus we come, through our own personal experiences, to the idea of a permanent and continuously moving *time*. But since we observe that these personal experiences are very intimately connected with events outside, and that often they are caused by such events, we attribute to external processes the ideas of earlier and later, and of shorter and longer durations. This succession of events in us, like the events outside, is called **time.** **Kant** believed that our idea of time was independent of every experience and that it was given before every event '*a priori*'. That was an unproved generalization. We could only say that, even without being conscious of these things, we simultaneously form with our first experiences the ideas of space and time. And as we can only speak of a length with reference to an object of which a beginning and end are somewhere apparent, so only can we speak of the duration of an event which, at some time, has a beginning and an end. It is true that we can speak quite intelligently about the duration of an unchanging state, but only if the commencement and the end are earmarked by some event or occurrence. We can no more speak of an everlasting invariable occurrence, than we can of an infinite length.

If we should now wish to inform someone how long an occurrence lasted it could be done in exactly the same way as was done previously for space. We then chose for the purpose of determining a length a body which was equal in length to that of the object being measured, whilst all its other characteristics could be disregarded. Nevertheless these other characteristics were so chosen that the body could be easily preserved and transported from place to place in order to be used there.

It now behoves us to seek a corresponding process the duration of which should be equal to that in question, and, moreover, the other properties of which should be quite immaterial. But as in the case of the unit of length, **this unit of time must be capable of being moved from place to place in time without thereby affecting its duration.**

Let us consider one example. In order to determine the time required for a sphere to roll down an inclined groove, that is to say, in order to be able at a later time to inform others about it, Galileo allowed water to flow out through a fine opening in the side of a large container, and weighed the quantity of water which had escaped during the time taken for the sphere to roll down, and which he had collected in a small vessel. In order to specify the duration of this event at a later time, to verify his results, or to compare them with other results from different places, he only needed to let the same quantity of water flow out from the same opening of the same vessel. He then said that the time taken for the water to flow out was exactly the same for the second event as for the first.

That is what he said—but was he really convinced of this ? Upon that everything depends. If we had asked him this question, how would he have replied ? He might have shown us that the head of water in his vessel was always the same, that the width of the orifice was invariable, and that he had been careful to use water which was always free from dust and sand. In other words, he would have shown us, that all circumstances which might affect the result in any way whatsoever had been carefully avoided, and he would be convinced that the same interval of time is always measured by the same amount of water which has flowed from the apparatus. If we expressed any further distrust he would probably have produced the sand clock of his day and demonstrated to us that, whenever the entire sand flowed once through the clock, a constant quantity of water flowed out from his vessel. We may have had further doubts. Perhaps the sand could have flown out at different rates. In that case Galileo would have replied that both the sand clock and the water vessel must have changed simultaneously and in the same manner. Perhaps he would have set several such sand and water clocks in action and have shown us that the indications of all these instruments always confirmed one another. After this we would have to assume that the times were really equal or that all these clocks had become faulty to the same extent. Afterwards some other factor which could influence these clocks in the same way might have been stated, but such a factor would be less likely to be contemplated, the more numerous and different in kind the clocks were.

Finally, therefore, there remains only our doubt as to whether time itself could exert any influence upon the clocks. In other words, we eventually arrive at the same limitation for the validity of our time indicators as we did when we were measuring a length. There we assumed that the mere shifting of the measuring rod in space did not alter its length ; we must now assume in a similar way that the mere repetition of an event in time does not alter the duration of the normal process. To time, as such, we can ascribe no influence on the processes of Nature capable of detection by the senses ; the same applies to place. But if anybody should again ask whether we could name any definite phenomena as a basis for this view, we should again reply, as we did for space, that there is no further ground for our assertion. In other words, what we measure with certainty, in the case of time, is only relative time.

The Unit of Time. Still more apparent than in the case of our measurement of length is the fact that it would be very troublesome if, after investigating the duration of some occurrence, one had to send to some other person who wished to be informed concerning this event, the quantity of sand or water together with the necessary instruments used to produce the flow of the sand or water. A similar way out of the difficulty has been found as in the case of our measurement of length. Now, in measuring a length, we found it necessary to divide that length into smaller lengths, each of which was equal to the unit of length. We now have to seek such a unit of time, with which all other times may be compared. At an early period sand and water clocks were used. After the substance had flowed through the clock, a simple inversion of the clock served to start it again. The turning around of the sand-clock corresponds to the shifting of the rod of unit length. In order to determine the duration of an event it was only necessary to specify the number which showed how often the normal event had been repeated during the process.

At a later date processes were discovered which permitted time to be determined more accurately, and much shorter durations of time to be measured. Above all, these discoveries possessed the great advantage that it was not necessary to restart the contrivance after only one observation. With much greater exactness and regularity than man alone could ever do the process continued quite naturally and without interruption. We refer to the motion of the pendulum, and of the balance wheel of a watch.

Just as the primary unit of length was derived from a great natural dimension, viz., the circumference of the earth, so the standard of time is that period which our earth requires for a single rotation around its axis, or more exactly, it is the mean time elapsing between successive transits of the sun across the meridian at any one place. This period is divided into 24 hours, every hour is divided into 60 minutes, and every minute into 60 seconds. Pendulums are generally designed so that their period of oscillation is exactly one second, and it so happens that such pendulums are about one metre long. The units of time in different countries were much less different than the units of length, even in the beginning, since the rotation of the earth about its axis, and the corresponding change to day and night, regulates the business life of this world quite naturally and uniformly.

In fundamentally the same way as for space and time, we shall have to estimate the magnitude of the different quantities which occur in the various branches of physics. It will always, therefore, be a question of finding a primary unit of the appropriate kind, which is either invariable, or which allows itself to be obtained again on every occasion. The discovery of such a unit, or the process whereby greater magnitudes of the same kind can be compared with it, is in no way the beginning of any branch of physical investigation, just as the adoption of the metre unit did not mark the commencement of the science of measuring length. On the contrary, the first laws to be discovered were expressed in preliminary and arbitrarily selected units, and it was only at a later period that it could be seen how best to choose the units, so that the laws could be expressed in a more coherent and suitable manner, by making the constants vanish wherever possible. The introduction of a final unit is rather the conclusion and marks the completion, rather than the beginning of any research into the fundamental laws concerning any kind of magnitude.

2. MOTION

Place and Time Coordinates. For quite a long time before we could define what we meant by motion, we were able to differentiate between a body in motion, and one which was not moving. A child sees the to-and-fro motion of a clock's pendulum, his attention being arrested whenever it ceases. Let us now ask ourselves, by what means we recognize whether a body is moving, and let us consider different motions in order to obtain a clear conception about this subject. We see a wanderer upon a country road vanish behind the different trees in turn and then appear again; or similarly, an express train rushing along through the fields, in contrast to one which is stationary, or a dog passing the houses as he runs along the street. In general, we are able to detect the motion of a body owing to the fact that its position, in relation to its surroundings, changes with time.

In order, therefore, to describe the motion, we must first be able to state the position of a body exactly, in order that any change which may come to pass may be detected afterwards. If we imagine, for a moment, that there were only one single body in all space—it can be indefinitely big and we ourselves might be part and parcel of it—are we then able to specify the position of this body? No. For the expression, it is here or there, has no meaning whatsoever, for there can be no 'here and there'. If we ever speak of 'here and there', this expression always implies the existence of neighbouring bodies, *i.e.* those objects which are in the vicinity of the one we are examining; they decide the meaning of the expression 'here and there'. The statement of the position of a body always signifies its spatial relationship with its surroundings. But it is not necessary to know the distance of a body from all the bodies in the world in order to fix its position in space with certainty. When its distance from three points, which do not lie in a straight line, is known, then its position is determined completely. For most purposes, three straight lines are chosen which are perpendicular to one another, and which intersect at one point. The three lines are called the coordinate axes, and the common point of intersection the origin of the coordinates. In order to define the position of a point uniquely we state how far it is from the origin, the distances being measured along these three axes, which are generally called the x, y and z axes respectively. The distances of any point expressed in terms of distances along each of these three axes are the coordinates of the point; they are denoted by (x, y, z). In order to find the position of a point in this system again, when the system itself, *i.e.* the origin and the directions of the three axes have been given and the three coordinates (x, y, z) of the point are likewise known, one proceeds, let us say, at first a distance x cm. from O along the x-axis, and arrives at A (Fig. 2); from A one goes y cm. in the y-direction, and arrives at B; from B one goes z cm. in the z-direction, so that P, the end of this last distance, is the desired point. It is the only point whose coordinates (x, y, z) have the prescribed values.

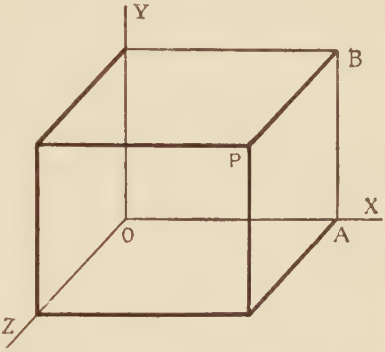


Fig. 2.

If it is now a question of events, *e.g.* of the arrival of a moving point at different positions in space, the time must also be known when the point arrives at different places.

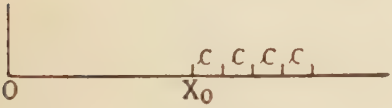


Fig. 3.

If the motion is not arbitrary, but takes place according to some definite rule, one is able by means of such coordinates to say where the moving body would be at any other time, without the necessity of constructing a long list, which, like a railway time-table, would state when the body should arrive at different places. Let us assume, for the sake of simplicity, a body which moves along the x-axis (Fig. 3). At the time when we commence to make our measurements, *i.e.* when $t = 0$, let the body be at x_0. Then if the body moves regularly in the direction of the x-axis, let it move a distance v in every second. Under these conditions only the x-coordinate can vary, and this increases v cm. every second. After t seconds, therefore, it has grown by an amount vt cm., so that the x-coordinate after time t is

$$x = x_0 + vt \qquad (1)$$

This equation permits us to state the position of the moving point at any time t, when x_0 and v are known.

It is only by means of space and time that we are informed of an event. If, therefore, the three coordinates (x, y, z,) which denote the position of an event in space are called the three space coordinates, the statement of the time ought, with just the same validity, to be added as a fourth coordinate. This idea, based on certain assumptions, was first carried out by **Minkowski** (1909), when he conceived the individual moments of time as points in a line which extends indefinitely into both past and future times.

In contrast to the space coordinates that which represents time possesses the characteristic that it can only increase. Time cannot go backwards. No body can recover again the time coordinate which it once had, whereas it may return to a position which it previously occupied.

Secondly, there is only one time coordinate, whereas there are three directions in space. Consequently, in order to specify the position of a body in space, three statements are necessary, whilst time is given by one statement. When the idea of 'Dimension', which hitherto was only used for the three directions in space, is extended to include that of time, **Minkowski's** theorem is arrived at, *viz.*, **every event takes place in a four dimensional space-time continuum.** Just as place is represented by a point in a three dimensional space continuum, so there belongs to every event a world-point, for whose complete determination four coordinates are necessary.

Since the space and time coordinates are continually changing when a body moves, the motion of a body is represented by a '**world-line**'. Two bodies meet when they arrive at the same place at the same instant. Under these circumstances, their two world-lines will either intersect, or touch one another tangentially, for all the four coordinates for both events are, at this moment, identical.

All this is quite intelligible, but one must not endeavour to form a mental picture of this four dimensional continuum. Since our space has three dimensions only, there is no direction available for the representation of time. But we can form a conception of the relationships if we contemplate a motion which occurs entirely in a plane. The direction which then remains free can be used as the time axis. For example, a point on a vibrating tuning fork, or violin string, moves to and fro between the points A and B upon the straight line OX, Fig. 4. After every swing the point is again at the same position in space, the time having advanced in the meanwhile. If we now wish to sketch the world-line, the direction OT can now be used as the axis of time. Since the body requires the same time for every swing, the time coordinate has increased by the same amount every time the point moves through its position of rest O. The world-line of the vibrating point is a wave-line.

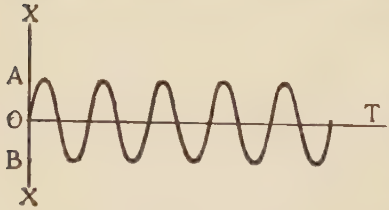


Fig. 4.

As a second example, we may suppose that a body moves uniformly along a circular path. Let this path be situated in the plane containing the y z-axes. Taking into consideration the constantly increasing time the spiral, Fig. 5, is obtained for the world-line, if the third direction is used as the time axis.

The following statements are now easily comprehended. The world-line of a fixed body is parallel to the axis of time. A motion represented by a line perpendicular to the time axis, in which all points would be reached at the same time, is a movement which takes place with infinite velocity. Between these two directions are distributed the world-lines of all motions having a finite velocity.

Such systems of coordinates, with their axes and origin, are not only used to represent motions in the laboratory of the physicist, but they are also used in business and by the private individual. In everyday life it frequently happens that we know the relations of a point to its surroundings so well, that we can designate its position simply by referring it to some known point in the neighbourhood. But it is here a question of discussing processes, which are always applicable, even when suitable objects are not present in the immediate vicinity. When we have to indicate the position of a ball upon a billiard table, it only remains for us to state its distance from each of two edges. The position of a small island in the ocean can also be stated explicitly, by saying, 'First travel from this point x km. east and then y km. north.' The designation of houses, in the large cities of America, as '15th street, house number 7', is nothing more than an expression of the x and y-coordinates of the house.

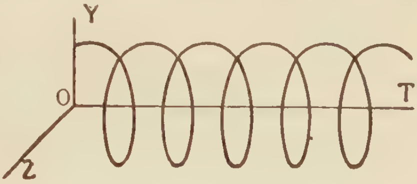


Fig. 5.

We have therefore established the fact that the position in which a body happens to be can be expressed by a statement of its position relative to its surroundings. And the motion of a body is known when we have information telling us how the position of the object with reference to its surroundings has changed, a system of coordinates being used for this purpose. In certain special problems the system of coordinates can be replaced by other equivalent variables. Thus, the inhabitant of a town might speak intelligently, not of his movement along the x-coordinate, but of his walk along William Street. But the determination of a point by means of its three coordinates is intelligible on all occasions without anything further being added.

Absolute and Relative Motion. It cannot be denied that anyone who is unbiased in his opinions will, at first, always maintain that our earth can be conceived as something fixed, so that every movement is completely described when its position, upon the earth, is specified by means of coordinates. It is referred to the earth as a coordinate system. As children we held the conviction that movements upon the earth represented the actual movements of objects. Just as with mankind as a whole, so every individual must live to see the time when he is convinced that our earth is not a fixed immovable object, but that it moves around its own axis once daily, and around the sun every year. The coordinate system upon which we describe our motion is in itself a moving system. What will now happen to the rules for specifying a position? Do they lose their significance, do they remain valid, or have they to be explained otherwise?

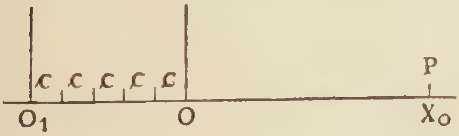


Fig. 6.

Let us assume that a point is determined by its three coordinates (x, y, z). But we now wish to assume that the system of coordinates itself, Fig. 6, moves in the direction of the x-axis, *i.e.* towards the left, with constant velocity v cm. per sec. Since the distance of the point P from the origin is varying continuously, its x-coordinate, which represents this distance, must undergo a similar change ; in fact it will increase by v cm. every second. The coordinate of any point P considered to be at rest will be, as it is usually expressed, a function of the time. If the coordinate has a value x_0 at the time $t = 0$, after a time t it will have the value $x = x_0 + vt$.

If this equation is now compared with equation (1), it is seen to be absolutely identical. We arrive at exactly the same equation whether we say that the point has moved towards the right along the x-axis, or whether the coordinate system upon which we have drawn the point has moved towards the left with the same velocity. If we now pause to reflect a little while, we are convinced that this ambiguity corresponds entirely with fact. We see the sun rise ; whether it is really the movement of the sun or the earth we cannot decide from the appearance of the event. From the beginning, as we ourselves do usually, mankind has held the primitive conception according to which the earth is a fixed point with the sun revolving around it. Only much later, for quite other reasons, has the opposite viewpoint been preferred. Every child will have had some experience similar to that which follows. When it first travelled by train it will have noticed that the telegraph poles at the side of the track appeared to glide along. Reference may also be made to the magic swing which is sometimes seen at fairs. You sit quietly in a hanging boat, which does not move, but on each side a painted landscape moves to and fro, the movement gradually being increased until you finally believe that you are hanging head downwards, and that you are afraid of falling, whereas you are really sitting in a fixed boat with absolute security. Such examples show us that we can really determine with our eyes when a body moves with respect to others, but we cannot decide from appearances which of the two has moved. We certainly cannot make such a decision when we ourselves take part in the movement of one of the two objects, and participate in all its motions. In other words, we are only cognizant of relative, not absolute motion. We have therefore to observe that the above considerations only refer to our perception of absolute motion from **appearances**, in other words, from the magnitude of the coordinates and their change with time. As to whether or not there are some properties of a moving body from which we could arrive at some conclusion concerning absolute motion, we shall have to speak later—when we discuss the effect of an existing motion upon the behaviour of bodies.

The Laws of Simple and Compound Motion. Every body can only move in one direction at any definite time, and, accordingly, there are only simple motions. But this simple motion can be compounded from several component motions. In reality, this always happens. In consequence of gravity all bodies fall perpendicularly to the ground. In a moving express train the falling body has, at the same time, the motion of the train, and likewise, all bodies move with the earth about its axis, and are carried by it around the sun at the same time.

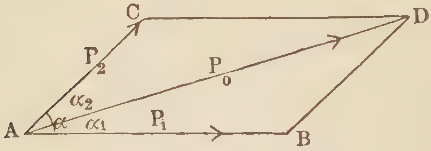


Fig. 7.

If a body gets from A to B in virtue of one movement, whilst a second would bring it from A to C, Fig. 7, when it participates in both movements at the same time, it will actually pass from A to D. A similar result is likewise obtained if a third or fourth component is added to these movements.

3. VELOCITY

We arrive at the idea of velocity by examining, let us say, the motion of two spheres which begin to roll together in the same direction towards the ground. The one hurries forward, traversing in the same time a greater distance than the other. Or we may consider two balls, which begin to roll across a table at the same instant, one of which falls over the edge sooner than the other. Or again, we may contemplate the hunter, mounted on a good steed, who overtakes the stag which he is chasing. If we could not make such observations, in other words, if there were only one kind of motion, so that every moving body travelled the same distance in the same interval of time, we should certainly not arrive at the idea of velocity. We should speak of the progress of bodies with this unchanging velocity as ‘ the ’ motion, without referring at all to their speed. But we now perceive that there are many degrees of movement ; the same body may move, in the same time interval, a greater distance at one time, and a smaller one at another. We say that the **velocity** of the object changes. Long before physics became an exact science, the idea of velocity, and its different magnitudes, was quite clear. In order to make our discussions as simple as possible in the beginning, we shall assume, for the present, that a body always retains the same velocity.

To physics belongs the task of discovering ways and means whereby the velocity of a body can be determined. Obviously, we must concede that two bodies have equal velocities, if at some time during their motion they are together, and they always retain this relationship afterwards, for they must travel the same distance in the same time. This definition may immediately be made general, so that, even if the movements do not take place at the same time, at the same place and in the same direction, velocities are said to be equal when equal distances are traversed in equal times.

But if a body travels twice (or n-times) the distance in the same time interval as another, it is said to have double (or n-times) the velocity of the first one. The same result is obtained if it traversed the same distance in half the time, for in the second half of the period it can travel once more the same distance ; it then has such a velocity that it could travel twice as far in the same interval. If we now have two bodies, one of which travels l cm. in t sec., and the other, l' cm. in t' sec., we cannot express an opinion about their velocities without proceeding further. It is first necessary for us to calculate how far each body moves in the same time, *e.g.* a second. Proceeding in this way we obtain the values l/t and l'/t' respectively. We therefore compare velocities by dividing the number of units of distance by the number of units of time, and then compare these values with one another. In the measurement of velocities we cannot proceed much further than this, for the matter is analogous to the comparison of two lengths.

The Unit of Velocity. It is now expedient for us here, once and for all, to select a definite unit of velocity, with which all other velocities can be compared. We are again quite free to choose whatever unit we like as the standard of velocity, only it must have the property that it can be used easily and reproduced with certainty. We therefore have to seek a motion, concerning which we can always say that equal distances will always be executed in the same time. If we reflect for one moment we shall soon see that it is not easy to find such a motion. Suppose that we finally adopt the following way out. The shadow of a church steeple moves at midday, between the hours of twelve and one o'clock, a definite distance across the churchyard. (The distance is not constant throughout all the days of the year, but this fact has no significance as far as we are concerned.) Let that be the unit of velocity. In order to compare other velocities with it, we must know how many centimetres this shadow moves in one second. Suppose we find it to be k cm. Then a body will have unit velocity when it moves k cm. per second. Consequently, a body which moves 1 cm. per second, has, according to this scheme, a velocity $\frac{1}{k}$, and if it moves l cm., the velocity is $l \cdot \frac{1}{k}$; if it moves l cm. in t seconds the velocity is $v = \frac{1}{k} \cdot \frac{l}{t}$. In all these expressions the factor $\frac{1}{k}$ occurs, and this helps to make the calculations wearisome. Now since any desired value can be chosen as the unit velocity, it would be much more convenient if k were unity, for we should simply have $v = \frac{l}{t}$, and v would be unity if $\frac{l}{t} = 1$. If, therefore, a body moves 1 cm. in 1 second it is moving with unit velocity.

Derived Units. This choice of a unit differs considerably from the methods which have hitherto been used. Up till now the unit of length has been chosen as the length of an arbitrary rod, and the unit of time, as the duration of an arbitrary event. We now set aside the idea of seeking a suitable velocity, which has to be realized as a definite state associated with a body, and base the unit of velocity on the units of length and time, which have already been fixed; whilst a *rule is specified* by means of which the unit of velocity is derived from the units of time and distance. The essential difference between this process and the former ones is as follows.

1. Hitherto the selected unit, *e.g.* that of length, has had to be fixed by means of a chosen body. One has to be careful that the length of this body shall remain as constant as possible, and that it can be very easily replaced if, at any time, it should be lost. In order to fix the unit of velocity by the new process we no longer need all this; if we have the units of length and time the unit of velocity is given at once by these.

2. But we have lost some of our freedom by so doing. Hitherto any length or time interval could be selected as a unit. This can no longer be done. As soon as the units of length and time have been chosen, the unit of velocity has been fixed and it is unalterable. If, for any reason, a different unit of length should be selected from the one which had formerly been adopted, *e.g.* one which is seven times as large, the unit of velocity would thereby become seven times as large, since a greater unit distance would be traversed in one second when a particle was travelling with unit velocity. But if the unit of time should be chosen to be sixty times as large, as when minutes are used instead of seconds, the unit of velocity would be thereby made sixty times as small, for now a body moving with unit velocity would have a time sixty times as long in which to run its course.

Thus the unit of velocity depends upon those of length and time in the following manner: it is directly proportional to the unit of distance and inversely proportional to the unit of time. This interrelationship between the units can be more shortly expressed by the following rule: The '**dimensions**' of a velocity are those of a length divided by a time; this can also be written in the form of an equation which expresses the relationship between the units themselves, and not, as is customary with equations, between the numerical magnitudes which are involved. For this purpose all symbols are either placed in square brackets or else printed with capital letters. Thus

$$v = [l/t] = [lt^{-1}] = LT^{-1} \quad . \quad . \quad . \quad (2)$$

This is called a dimensional equation.

On this account the units of length and time are called **absolute units,** or more correctly **independent units,** whereas velocity, and other similarly derived units, with which we shall later have an opportunity of becoming acquainted, are called **derived units.**

Besides the units of length and time, there is only one other independent unit, viz., the unit of mass, and that is the gram. All the remaining units are based upon these three fundamental units. For this reason our system of units is called the Centimetre–Gram–Second (C.G.S.) system, or the absolute system of units. In England the foot, pound, and second are the three fundamental units; this system of units is known as the F.P.S. system.

For all derived units there is therefore a dimensional equation which allows us to recognize, at once, the manner in which the different units depend upon the fundamental ones. Strictly speaking, only the three symbols m, l, and t, or powers of these, can occur; the occurrence of these in the denominator is expressed by means of a negative power, -1, -2, . . .

Let us now derive a few dimensional formulæ in order to study these ideas more fully. We are accustomed to consider the square of 1 cm. length and breadth as the unit of area. But that is not absolutely necessary, for we could cut out, let us say, a rectangular piece of lead, and define its surface as the unit of area. Then one has to discover how many such pieces would be contained in a square of 1 cm. length and breadth. Let us assume that there would be k pieces, so that 1 cm.^{2} would be equivalent to k units of area. Then a rectangular surface of length a cm. and breadth b cm. would contain $k \, . \, a \, . \, b$ units of area (S). The value of k, which now occurs in all expressions of area, depends upon the unit of area which has been selected. If this is so chosen that the square of 1 cm. side is considered as the unit of area, then is $k = 1$, and we have simply $S = a \, . \, b$; familiarity with this equation makes us consider it as being self-apparent. But the unit of area is thus dependent upon the unit of length, and if a unit of length three times as big should be introduced, a new unit of area would immediately be defined, having a length and breadth which were increased threefold, so that it would be $3^2 = 9$ times as large as previously. The dimensional formula for an area is therefore $S = [l^2]$. Likewise, volume has dimensions $V = [l^3]$.

One must proceed in the same manner in order to arrive at dimensional expressions for all other units. If, in the course of research, a new idea is encountered, a numerical relationship between this new magnitude and others already known is first sought. Frequently an arbitrary magnitude is chosen for the unit, and the relationships which have been discovered are expressed by means of a mathematical formula in which, except for the new variable, and the others which are already known, only a constant occurs, the magnitude of which depends upon that of the newly selected unit. The new unit is then chosen so that this constant is unity. The formula gives at once the dimensions of the new unit.

But it does not always happen that the formulæ are found so simply as in the above example, by defining the unit of area on the basis that an area is measured as the product of a length and a breadth. Later on it is not so easy. The exact laws are only obtained by means of the most careful physical investigations, and it is only when the relationship has been obtained, quite free from doubt, and with the highest attainable accuracy, that one may proceed to fix the new unit. The choice of an absolute system of units could not possibly form the starting point for research, but must rather be regarded as the conclusion, and the crowning of the investigations concerning the quantitative relationships which arise in any such work.

Variable Velocities. Until now, in order to simplify our considerations as much as possible, we have assumed that the velocity of the moving particle remained unchanged. In the language of mathematics we may say that, whether a great or small distance l is chosen, and the corresponding time t, in which it is traversed, measured, one always obtains the same value for $\frac{l}{t}$. Conversely, from the given velocity v, and the time t, the described path l may be derived, thus: $l = vt$.

Actual motions are seldom, if ever, so simple. A falling stone or a departing train are easily seen to move less quickly at the beginning than at a later period of their motion. A horse becomes tired after drawing a waggon and goes more slowly. A clock's pendulum changes its velocity continuously.

It is now easily seen that, even when the velocity of a moving body is no longer constant, *e.g.* if a ball which is rolling uniformly along a table comes to an inclined plane where it goes quicker and quicker, the idea of velocity for this ball is not without meaning. Even now the ball has at every instant a definite velocity even although this is continuously increasing; but, on the other hand, it is also clear that the velocity can no longer be ascertained, as hitherto, by choosing a length of path $x_1 - x_0$ and determining the time $t_1 - t_0$ in which it is traversed. By calculating the value of the expression $\frac{x_1 - x_0}{t_1 - t_0}$ the constant velocity is obtained which a body must possess in order to traverse the distance $x_1 - x_0$ in the time $t_1 - t_0$. But the velocity of the rolling body cannot be permanently greater than this value, for then it would move a greater distance than $x_1 - x_0$. Neither can it be permanently less, for then it would move over a distance which was too small. Its velocity, which according to our hypothesis is not constant, must therefore be partly greater and partly less than this value, which represents for an irregularly moving body the mean value of its velocity. Moreover, it is clear that its true velocity at different points in its path, or at different times, is different from this average value. The value of the velocity which has been found must therefore be intimately connected with the distance and time; it will be a function of the time and of the distance.

In order to obtain a simple statement regarding the meaning of velocity when the latter is variable, the following procedure may be adopted. It can be assumed that, at the specific instant when the velocity is to be measured, the velocity suddenly ceases to vary, so that the body now advances uniformly along a distance with just that velocity. The velocity, which is now uniform, can be measured in one of the ways which have been described above. The result of such measurements is the velocity which the body actually had at the instant when it ceased to vary. There are special cases in which this imaginary experiment can actually be performed, and used to determine the velocity we refer to as in Galileo's inclined groove, or **Atwood's** machine.

But, in general, this cannot be done for all motions ; as instances of such a motion one may mention the to-and-fro motion of a pendulum, or tuning-fork.

The object, however, can be attained in the following way. Since, according to the laws of mechanics, which are soon to be discussed, the velocity of a moving body cannot change by fits and starts, the mean value of the velocity obtained by measuring a distance and the corresponding time interval will deviate less from the actual value at the first instant, the smaller the magnitude of the time interval involved. If the motion is specified by a mathematical formula, it can be proved that the expression $\frac{x_1 - x_0}{t_1 - t_0}$ approaches a finite limiting value more closely, the smaller the time $t_1 - t_0$ is made. It may be found, for example, when the motion takes place under gravity, that the path described is proportional to the square of the time; thus

$$x_1 = ct_1^2 \text{ and } x_0 = ct_0^2.$$

From this it follows that

$$x_1 - x_0 = c\ (t_1^2 - t_0^2) = c\ (t_1 + t_0)\ (t_1 - t_0).$$

Hence $$\frac{x_1 - x_0}{t_1 - t_0} = c\ (t_1 + t_0).$$

The more closely t_1 approaches t_0, the more nearly does this expression become $2\ ct_0$. The limiting value, which is reached when $t_1 = t_0$, is called, in mathematics, the differential coefficient of the path x with respect to the time t. This is written in short :

$$\lim_{t_1 \to t_0} \frac{x_1 - x_0}{t_1 - t_0} = \frac{dx}{dt} = v_0 \quad . \quad . \quad . \quad . \quad (3)$$

Whenever the motion is variable, the velocity is found by dividing the short distance, which is traversed in any short interval of time, by the time in which the path is traversed. If this fraction approaches a limiting value when this time interval is made smaller, this limiting value is the velocity of the particle at the instant which is under consideration. In particular, we have seen from the above example, where the distances traversed were proportional to the square of the time, that the velocity must increase regularly with time, *i.e.* the velocity must increase by the same amount ($2ct_0$) every second.

4. ACCELERATION

A motion in which the velocity changes is called an accelerated motion, a retardation being conceived as a negative acceleration. An acceleration is greater the greater the change in velocity and the smaller the time in which this change takes place. If the velocity increases by the same amount in equal intervals of time the acceleration is said to be uniform, and it is measured by the increase in velocity per unit time. If we take as the unit of acceleration, that acceleration which causes the velocity to increase or to diminish by unit amount per second, then the change in velocity Δv, in the time Δt, gives us the acceleration, *i.e.* $a = \frac{\Delta v}{\Delta t}$.

In the above example where the distance traversed was proportional to the square of the time, we found for the velocity the expression $v = 2ct$. It increases by the same amount every second. Such a motion has been called a uniformly accelerated motion, and the acceleration, or increase in velocity per unit time, is equal to $2c$. Conversely we could conclude that, when a body moves so that the acceleration is uniform, *i.e.* $a =$ const., the velocity is expressed by the equation $v = at$, whilst $s = \frac{1}{2} at^2$ expresses the path traversed.

In the same way, as when a non-uniform velocity was under consideration, if the acceleration is not uniform a mean value must be chosen ; the value thereby obtained will approach the true value for the acceleration more nearly the shorter the time of observation becomes. The limiting value, when $\Delta t = 0$, is the true acceleration at that instant. We may therefore write

$$\lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t} = \frac{dv}{dt} = a \quad . \quad . \quad . \quad . \quad (4)$$

The dimensions of an acceleration are

$$a = [vt^{-1}] = [lt^{-2}] \quad . \quad . \quad . \quad . \quad (5)$$

In general, a is not a constant, but depends upon the law which governs the motion.

Since observations concerning changes in position must always be made relative to some system of coordinates, it follows that the values derived for the velocity and acceleration can only be relative values.

When we were discussing velocity, it was observed, that any velocity could be considered as the resultant of two or more component velocities. Similar remarks may be applied to accelerations, for any acceleration may be regarded as the resultant of several component accelerations.

CHAPTER II

THE MOTION OF BODIES

1. The Teachings of the Ancients concerning Motion

NOW that we have formed a very clear conception concerning the general nature of motion, we wish to become acquainted with the laws which govern the motions occurring in the universe. Just as motion was certainly the first natural phenomenon to be noticed by mankind, so was the attempt to understand this process certainly the beginning of natural science. In fact, we have no theory of any other natural phenomenon which has been so completely worked out, and which goes back so far into the grey antiquity of time, as does the theory of motion. There is hardly any other subject which is able to show so clearly the difference between the methods which present-day scientists use to investigate a problem, and those which were advocated by the ancient Greek philosophers. Moreover this Aristotelian teaching was the cause of a long conflict between the ancient and modern schools of thought which marks the beginning of the modern era in science; it will therefore repay us to consider, somewhat in detail, the treatment which this question received in ancient and modern times.

Let us first of all direct our attention to the conceptions which the old Greek school of philosophers held concerning the material world.

Aristotle had declined to accept the atomic hypothesis of **Democritus** and had applied himself to study the older teachings of Empedocles concerning the elements. Earth, water, air, and fire, are the four elements from which all earthly bodies are formed. For the heavenly bodies a special substance—the ' quinta essentia '—called the Æther was postulated. *A priori*, **Aristotle** believed that he was able to prove that there could only be four elements because every material substance must contain one of the two opposites—warm–cold and moist–dry. From his teacher, **Plato**, he had received the doctrine of natural places: the earth is heavy, it endeavours to move downwards, there it belongs and its nature is such that, having been removed by violence from its original position, it tends to return to that position. Fire is light, that is to say, according to Aristotle, it has its place above, and thither it tends to go. Water is relatively heavy, arranging itself above solids, whilst above it is the place for gaseous matter. The paths of the heavenly bodies, which are formed from that supramundane fire, the *Æther*, are also situated in this region of fire.

Aristotle was convinced that the nature of things could be learnt from their behaviour, since this was a result of their nature, which must therefore reveal itself in this behaviour. The most general of all phenomena are those appertaining to motion. Now there are some motions which only arise because some powerful external force has constrained them to leave their natural positions. As an example of such a motion we may cite a stone which has been thrown upwards. From such movements nothing can be learned concerning the nature of bodies, simply because they do not arise naturally, but have been occasioned by outside influences. But there are also other movements which the objects carry out of **their own accord**, and it is from such movements as these that the inner nature of the objects will be revealed. We observe that solid bodies tend to fall downwards, that they seek the lowest position in water, that gas under water tends to rise, and that fire endeavours to rise above the air. These motions are therefore determined by **the nature** of the body concerned, and need no external cause.

If we attempt a resumé, we may say that there are four **essentially** different elements which appear in quite different forms, as solids, liquids, gases, and fire, and to each of these a definite place in the universe has been assigned. By their very nature they desire to be in their own assigned place, and when they are removed from it by force, it is this nature which sets the body in motion.

A certain logical accuracy cannot be denied to this view. It seems, moreover, that when experience has been appealed to, the result has been in favour of these assertions. However, if the hypothesis is examined more closely, it is found to be deficient in many ways. We are quite certain that the three states of aggregation do exhibit many marked differences, but the idea that these are differences in their essential nature was not the result of observations,—it was arbitrarily assumed by the observer.

The whole further development of the study of motion is based upon this statement. Indeed, some facts (the naturally occurring motions) appeared to confirm this hypothesis, but this confirmation is only forthcoming, if, from the circumstance that no external agent to cause the motion could be specified, the conclusion is made that no such agent exists, and that the bodies move in virtue of some influence inside them. These ancient philosophers did not consider the fact that their doctrines established a great gulf between motions which took place naturally, and those which were occasioned by some external agent, whilst Nature herself provides us with an example of a motion, viz., that in a horizontal direction, which is quite a balanced transition from the one to the other variety. Aristotle knew these facts quite well, but he passed too quickly from facts to theory. On one occasion this philosopher stated that by observing natural occurrences continuously, one acquired the facility to build hypotheses which would correlate whole series of facts with one another, but if ideas are studied for too long a period, this facility was impaired : we became alienated from the appearance of the real thing, for we tended only to pay heed to a limited number of facts. When the field of view is so restricted, there is a tendency for untenable hypotheses to be developed. It appears hardly possible to characterize the researches of Aristotle more fittingly than he has himself done above (Gomperz I 256). Also in his ‘ natural philosophy ’ Aristotle worked primarily with ideas.

2. The Laws of Motion according to Galileo and Newton

Modern teaching concerning motion makes no kind of assumption concerning the nature of the body, nor the different types of motion. It commences much more simply with experiments ; in these experiments one examines the manner in which bodies behave under the different conditions which we know exert an influence upon the motion. We have experienced in our arm the sensation of a certain resistance when we endeavour to set a body in motion, even when the body is not fixed to its position of rest, as for example, a leaden sphere which hangs by a cord from the ceiling: or we may say that a certain effort has to be made by us in order to set the body moving. Moreover, it has been discovered that certain natural contrivances exert exactly the same effect upon bodies as our arm does when we feel an exertion in it: thus, a stretched spring, a falling mass, etc., can set bodies in motion.

Force and Motion. 1. We can easily convince ourselves of the fact that the weight of a body is not changed to any observable extent when a body is displaced to a slightly higher or lower position. For this purpose the following apparatus is erected. The leaden sphere *K*, Fig. 8, whose mass is approximately 1 kgrm., can move with very little friction along the rails SS_1, from which it is supported by means of wheels, so that it can move with very little friction. The rails are slightly inclined, so that the carriage, when it is carefully adjusted, remains stationary everywhere, but moves without appreciable acceleration or retardation when it receives a slight push towards one end of the rails, since the frictional forces have been almost entirely compensated. The carriage may be caused to move by means of the small weight *W*. This little weight operates upon the carriage through the agency of the cord *F* which passes over a light pulley-wheel. Since the tractive force of *W* is independent of its height, it acts as a known constant force upon the carriage, whilst all other influences have been eliminated. If a metronome which beats seconds aloud is set in motion, and the carriage is set free at a moment when a beat is heard, then the position of the carriage 1, 2, 3, . . . seconds after the commencement of the experiment can be observed. The observations may be as follows:

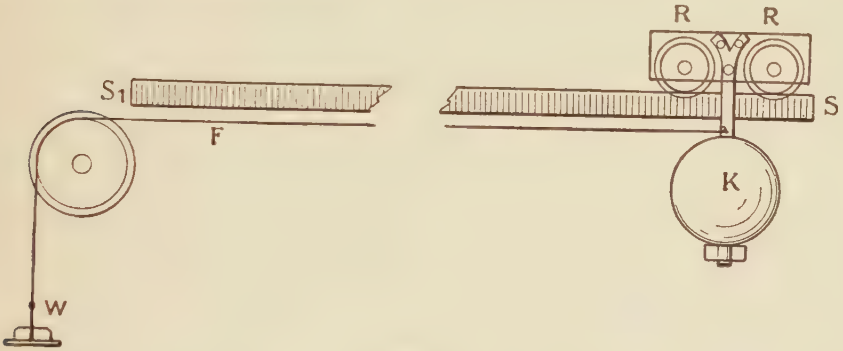


FIG. 8.

TABLE 1

Time in seconds	1	2	3	4
Distance traversed (cm.) . .	9	36	81	144
or . .	1 × 9	4 × 9	9 × 9	16 × 9
or . .	1 × 9	$2^2 \times 9$	$3^2 \times 9$	$4^2 \times 9$

Expressed in words, the above table shows, that the distance traversed is proportional to the square of the time, and we know that this only occurs when the motion is uniformly accelerated.

However often the experiment is repeated the same result is always obtained. We may modify the experiment somewhat, by using a larger or a smaller mass W, when the velocity will be different, but the motion will always remain uniformly accelerated. The experiment can be performed in other localities, or the direction of the rails altered, or we can perform the experiment by night, in summer, or in winter, in order to collect evidence concerning the possible effect which the time of day may exert upon it. (One remembers the varying position of the earth with respect to the stars.) All these conditions are found to be without influence, for the motion is always uniformly accelerated, and the acceleration is only changed if either the moving body or the suspended mass is varied.

Our experiments therefore corroborate the law that:

All bodies are uniformly accelerated when a constant force acts upon them.

The important point contained in this result is as follows. At first, the force acts upon a stationary object, thereby imparting to it a certain increase in velocity per second. Later, this same force acts upon a moving body, and still later, upon a body which is moving faster and faster, but the increase in the velocity per unit time is always the same. **The acceleration is independent of the velocities which already exist in the direction of the motion,** *i.e.* the acceleration, which a given force produces in a body, is not affected by the velocity of that body.

2. We are now able to obtain two equal forces. If a second small mass is secured, the material, size, and shape, being absolutely identical with that of the first body, we can verify the fact that it would give to all bodies the same acceleration as the first. When both are therefore suspended from the cord at the same time, the carriage experiences a force equal to twice its original magnitude. If the path which it traversed is again determined, as above, it will always be found to be twice as great as in the first case. Likewise, when the force is increased 3, 4, or more times, the path traversed in the same time is found to be 3, 4, or more times larger, than when the single force acted alone.

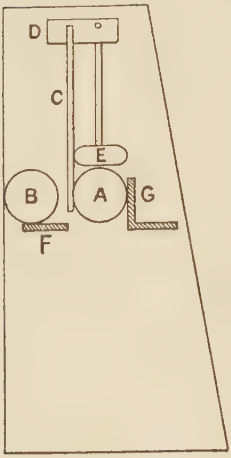


Fig. 9.

This result may be explained again as follows. The first force acts upon a body upon which no other force is operative, and imparts to it a certain acceleration. The second, third, and fourth forces, etc., act upon a body which has already succumbed to the action of one force, but yet they impart the same acceleration as the first. We may therefore say that the **acceleration produced by a force acting on any body is independent of the accelerations which other forces may be creating in it at the same time.**

3. There still remains another question: How does a body behave with respect to an accelerating force when, at the same time, another force is operative in another direction? This question can only be answered experimentally. If two spheres, of the same material and size, are allowed to fall at the same instant upon the surface of a horizontal floor, only the noise of one impact is heard. In the apparatus of Fig. 9 let A and B be two equal spheres. A is held in position by means of the piece of wood C, whereas B lies upon a small board. If C is struck by means of the

hammer E, A is immediately set free and falls in a straight line to the floor, but the sphere B is projected horizontally by this same blow. Thus there is associated with the falling motion of B one which takes place in a horizontal direction. But this last component causes the body to move away from the upright, and impact upon the ground only occurs when a distance, measured vertically, has been traversed, which is equal to the distance through which A falls. If this distance is traversed in the same time in spite of the horizontal component of motion only one impact will be heard. Experiment decides in favour of this hypothesis. It is thus found that a motion is not upset by the fact that the object experiences another motion in another direction at the same time. In other words, **the movement of a body in one direction is not affected by a movement in any other direction.**

This behaviour of bodies is referred to the principle of the independence of motions, and it is of very great significance. The most important consequences are here recorded.

1. If a body possesses several motions at the same time, it arrives at the same place as it would do if the motions had simply been added together one after another. This place can therefore be found by calculation or by drawing the parallelogram which is representative of the motion, as we have already done on another occasion [*cf.* Fig. 7.] There is, however, an essentially different point of view to be noted. **There we assumed** that the object made two movements, **and inquired as to the final position of the body**—a purely mathematical problem. We now ask the question : Do terrestrial bodies make both motions without disturbance ? The question is now a physical one covering the behaviour of bodies in our world. The answer which experiment gave to this question was that motions do not mutually disturb one another, and it follows that we can apply to the bodies of our world the method previously applied to find the position of the body.

2. Since forces are measured by the acceleration which they impart to bodies, these accelerations can be used to represent their magnitude. And since bodies can follow the combined efforts of several forces simultaneously, the resultant of several forces can be found from the parallelogram rule. The resultant is that force which, acting alone, would have imparted to the body the same acceleration as the two, three, or more forces do together. But we must be careful not to deduce more from this law than has been established. It has only been proved that this resultant force can replace two single forces in respect of the acceleration which the body experiences, and not in respect of the internal stresses which might be caused by the component forces. For example, two equal forces acting in opposite directions would produce zero acceleration, but since they act in contrary directions they would endeavour to tear the body apart, or, if it were not sufficiently rigid, it would actually be torn asunder, or crushed to pieces.

3. Since the behaviour of bodies towards any one force does not vary in any way with any incidental motions which the body may have in the same or in any other direction, we cannot conclude from the behaviour of the body whether or not any such movements exist. In a moving express train a sphere will fall towards the floor in a direction which appears to the traveller to be vertical, just as it would do outside the train, so that the sphere can give no information as to whether the train is moving or at rest. If, therefore, we had learned from visual observation that

there is a relative motion between the sun and the earth, we should not be able, from the behaviour of these bodies and their movements, to say whether the earth or the sun was moving. From this principle of independence there results, however, a very important extension which must be added to what we have already said concerning absolute and relative motion.

Everything that has hitherto been said concerning the motion of bodies can be summarized in one sentence : the acceleration produced by a force is independent of the body's state of motion, and depends only upon the magnitude and direction of the force.

Mass and Motion. Another question still remains to be answered : Is the acceleration imparted to a body independent of the body which is moved ? In order to investigate this problem several bodies identical in material, size, and shape, are first procured, so that we can assume, with absolute certainty, that their behaviour towards any force, tending to produce motion, shall be the same. A suitably chosen experiment will confirm this assumption. The same force is allowed to act upon two bodies at the same time, and the acceleration measured as on previous occasions. It is found to be half as large as before ; if a three- or four-fold quantity of the same material had been used, the acceleration would have been found to be three or four times smaller, etc. We find, therefore, that with bodies, such that for which the quantity of material in them can be determined from their volume, the acceleration is inversely proportional to the quantity of material in the body. From this it is obvious that bodies do influence the motion quite definitely, and that the acceleration becomes less the greater the quantity of matter which is set in motion. That property of bodies, which determines the acceleration which a given force may impart to them, is called **mass**. Now since it has been found that, for substances of the same material, the acceleration is inversely proportional to their masses, different substances can be said to have the same mass if they experience the same acceleration when acted upon by the same force ; moreover, if the accelerations are different, then the masses are inversely proportional to them.

Now that we have discovered that property of bodies which determines whether they will be differently accelerated by the same force, we have created a new idea, viz., that of mass. We have already seen under what circumstances two masses can be said to be equal, and how we can compare different masses with one another.

The unit of mass has been arbitrarily fixed as the mass of a cubic centimetre of pure water at 4^0 *C*, *i.e.* it is independent of other units. It is called the *gram* (*grm.*) The mass of a cubicdecimetre or one litre of water is 1,000 grm. and is called the kilogram (*kgrm.*) Since the exact production of such a quantity of water is difficult and troublesome, a standard kilogram mass of platinum-iridium, an alloy which is not corroded by the atmosphere, has been made, and it now forms the practical standard of mass with which all others can be compared.

We now wish to try and ascertain, more exactly, what it is in bodies which conditions their different behaviour with respect to the same force. The manner in which this may be solved is as follows : one by one the characteristics of the moving body are altered, and every time we observe whether the acceleration is thereby affected. If we first of all investigated the influence of such incidental properties as shape, colour, and temperature, etc., we should find that these did not affect the acceleration.

We proceed at once to examine the more inherent properties of matter. It is very striking to observe that the behaviour of a body in respect to its motions is not affected by its state of aggregation. One might consider the views of the ancients about motion. In order to verify the above remarks we may hang from the carriage of our suspended railway a light hollow vessel and, having introduced a piece of ice into it, the acceleration due to any arbitrary force may be determined as before. The ice is now allowed to melt, without any being lost, and the acceleration produced by the same force is again determined—it is found to be unaltered, a fact which proves that the state of aggregation is without influence upon the acceleration. The same result would be obtained if we procured an apparatus in which the body could be allowed to pass from the liquid to the gaseous state.

The same result is obtained if a salt is introduced into the carriage, first separated from the water and afterwards dissolved in it. The acceleration is constant.

There still remain, however, the changes due to chemical processes: these changes are such that, when they do take place, the reacting substances are considerably altered. Let us assume that the carriage contains substances capable of reacting chemically with one another. As an example we may cite the case of aqueous solutions of silver nitrate and sodium chloride, which are provisionally separated from one another. Suppose that the acceleration is a under the influence of some force. The two solutions are now mixed whereby a chemical action ensues. Silver chloride, separating out as a solid precipitate, and sodium nitrate in aqueous solution, are the products of the reaction. If the same force is again allowed to act upon the system the acceleration is still found to be a exactly. Chemical actions, therefore, do not affect the mass of the participating constituents.

Mass is so inseparably associated with bodies, and so unalterable in its magnitude, that it can never be separated from them; neither may it be increased nor diminished. Now everything in the inanimate world consists of events, either the result of purely mechanical processes, or physical changes (produced by heat, melting, evaporation, or solution), or chemical changes, which give rise to new substances. Since, however, we have seen that the mass of the participating bodies remains unaltered under all these conditions, it cannot, in general, change. In any given closed space the mass can only be increased by introducing new matter.

If, by a closed space, one is to understand a space into which or from which matter cannot be removed, the following law holds: **In any closed space the mass is constant.**

If the universe is finite, then it can be considered as such a closed region, and the law may be expressed as follows. The mass of the universe is constant. If, however, anyone held the view that the universe was infinite in extent, there could be no possible meaning in the statement that its mass did not change, for a change in something infinite cannot be demonstrated.

The statement concerning the unalterability of mass can equally well be expressed as follows: **Mass is an additive property.** That is to say, if one body is added to another with which it combines physically (as when it passes into solution), or chemically, it still retains its own mass, which is added to that of the other body, so that the mass of the whole is equal to the sum of the masses of the constituents. Other characteristics of matter, such as volume, are not necessarily additive. For example, if a litre of water is mixed with a litre of alcohol, there are no longer two, but 1·9 litres.

The relationships which have been discovered may be summarized as follows :

1. If the same body is accelerated by different forces, f and f', the accelerations, a and a', are proportional to these forces.

If m is constant, then $f : f' = a : a'$.

2. If different masses m and m' are accelerated by the same force f, then the resulting accelerations are inversely proportional to the masses.

If f is constant, then $m : m' = a' : a$.

3. From these two laws another may be derived which may also be verified by another experiment.

If, by increasing the force, the acceleration of the larger mass is made equal to that of the smaller, then the two forces must be related to each other in the ratio of the two masses.

If a is constant, then $f : f' = m : m'$.

All these relationships can be expressed by means of one law. Let us assume that, for the time being, some one force is chosen as a unit, and it has been found that C such units are necessary to impart unit acceleration to unit mass. The magnitude of C will depend upon that of the unit which has been selected. Then, according to the first law, a force which is a times as large as this unit must be used in order to give a mass of one gram this acceleration a. But should a m-times bigger mass have this same acceleration, so must the force be still m-times increased. We therefore have

$$f = Cma \quad . \quad . \quad . \quad . \quad . \quad (6)$$

By the choice of a suitable unit of force, C is made as simple as possible ; preferably, it is made equal to unity. If that is so, the force which gives unit acceleration to unit mass has been chosen as our unit of force. This force is called the *dyne*. The unit of force has the following dimensions :

$$f = [ma] = [mlt^{-2}] \quad . \quad . \quad . \quad . \quad (7)$$

and the final law concerning force is

$$f = ma \quad . \quad . \quad . \quad . \quad . \quad (8)$$

In this extraordinarily simple law, all our teachings concerning motion are embodied. It is valid, without exception, and without restriction, for all motions of all bodies.

So long as there is no force ($f = 0$) acting upon a body the acceleration is, according to this law, zero, and the body moves without being accelerated. Should, however, the acceleration a become infinitely large, that is to say, should the velocity become infinitely big in an infinitely short time, then an infinitely great force would be necessary.

If we ask, finally, how far this law has been proved, the reply must be as follows :

1. The law is proved by direct experiment. Naturally such experiments can only be made with bodies on the earth which can be accommodated in the laboratory. Inside these limits the law has been verified. Expressed in a rather better way, we may say that the more we have been successful in removing those extraneous forces, which may influence the result, the more closely has the above law been confirmed.

2. On the basis of these experiments the law has been assumed to be valid in general. It has been used in calculations, and conclusions based thereon have been utilized in the construction of machines, flywheels,

railways and motor-cars; in short, it has been applied wherever motion enters into the problem, and never has it been found to be at variance with experiment. The modern development of industry and technology is a very convincing, but indirect, proof of this law.

3. For the heavenly bodies every direct proof is certainly excluded. Nevertheless, this new law of motion was applied, even in the beginning, for the purposes of astronomical calculations. By means of this law one can predict, for example, the time and place of a solar eclipse, which is to occur a hundred years hence. Such predictions have been made repeatedly, and never has any discrepancy been detected. More recently spectrum analysis has shown that the heavenly bodies are composed of the same elements as are found on this planet, so that we must assume the same laws of motion for them if we do not wish to be guilty of the greatest caprice.

Newton's Principle of Relativity. With regard to the manner of presenting and expressing things, as is now customary in the more recent theory of relativity, we shall once again consider the above in detail, but from another viewpoint. In the remarks which follow we shall therefore adopt those expressions which are used in the Theory of Relativity. We shall begin these deliberations with an example which played a part in the history of physics at the time of Galileo. It concerns the following question: Does a sphere, which falls down from the mast of a moving ship, strike the deck at the foot of the mast, or does it, as was then generally believed, strike the deck at a spot which is just as much behind the mast as the ship has travelled forward in the meanwhile? When the experiment was made it showed that the sphere always fell against the foot of the mast, irrespective of its height or the ship's speed. If it had been otherwise, the speed of the ship could have been ascertained from this phenomenon. Then it would have gone during the time of fall just as far as the sphere had fallen behind, but since the sphere did not remain behind this means of ascertaining the velocity was excluded. The whole process can be represented thus: an observer travelling with the ship would observe that the sphere fell perpendicularly to the deck. We shall soon see that a falling motion is a uniformly accelerated one, so that it can be represented by the equation, $z = \frac{1}{2} at^2$, if the vertical direction is represented by z. If the ship itself is now moving, and the true motion of the sphere is to be specified, the motion of the ship must be considered. Let us assume that the ship moves with velocity c in the x-direction, and that this velocity is constant during the short time required for the body to fall. Under these circumstances the motion of the sphere is compounded of two motions, one directed vertically downwards, and the other in the x-direction. If a traveller on another ship lying close by, and at rest, wished to fix the place of the sphere at different times, he would actually use the two equations, $x = ct$, and $z = \frac{1}{2} at^2$.

Let us now assume that, on the following day, the two ships pass one another again, but this time the first ship is stationary and the sphere, which falls from the mast, has, in reality, only this one motion. The other ship, however, moves in exactly the opposite direction with this same velocity c. If now the person on the second ship again observes the co-ordinates of the falling sphere he will again perceive the changing z-co-ordinate, in fact, it will be just the same as he found yesterday, viz., $z = \frac{1}{2} at^2$. This equation is not deduced by any mathematical reasoning, but from the experimentally established law concerning the independence of motions. The falling motion is exactly the same to-day, when the ship is at rest, as it was yesterday, when the ship was moving. But, in addition, the observer will to-day be able to establish a change in the x-coordinate, not because the first ship is moving, but because his ship containing the origin of the coordinates is in motion. Since now his ship is moving exactly in the opposite direction with the velocity c, the x-coordinate of the sphere is therefore $x = ct$, exactly the same as yesterday.

And if the falling experiment is continued on a third day when the ships are again passing by, there being no fact from which one could tell which ship is moving, but only the relative motion of one ship with respect to the other is seen to change, it would not be possible to discover which ship had moved because the same result is always obtained. Actually the event (the falling sphere) has been referred, in the first instance, to a system of coordinates at rest (the stationary ship), whilst a moving system has been used in the second case. The acceleration of the falling body has been identical in both instances. It follows, therefore, from the law of the independence of motions, that the law governing the motion of the falling body is always unchanged, whether it is referred to a fixed system or to one which is moving. If the reference system of coordinates is moving, say, in the x-direction, then it is only this coordinate which will be different; in fact, if c is the velocity of the coordinate system the x-coordinate varies according to the equation $x' = x + ct$; the y- and z-coordinates do not vary.

For dynamical processes we therefore have the theorem, that the laws of dynamics remain unchanged whether they are referred to a stationary, or to a moving, system of coordinates. This is called the **Newtonian Principle of Relativity**, and the very phrase itself implies that only relative motions can be recognized. If the laws had taken another form according to the system of coordinates used, amongst all these different forms there would have been one which, perhaps by its simplicity, stood apart from the others—this would then have been the system of coordinates which was at rest, and motions referred to this system would have been characterized as absolute motions.

There are an infinite number of systems which, for stating the laws of mechanics, are all equivalent.

The principal law of dynamics states that a body always moves with the same constant velocity along a straight line, if it is not acted upon by any forces. Strictly speaking, the law, expressed in this form, is quite unintelligible, for the exact sense of a line can only be recognized when it is referred to a system of coordinates. The motion of the sphere which fell from the top of the mast appeared from observation to be that of a straight line, but actually it is most certainly not so, for the falling body has, at the same time, participated in the earth's rotation about its axis, its path being therefore curved. We must therefore always state exactly the system of coordinates in which the line is said to be straight. Now it is mathematically true that all lines which are straight for one system of coordinates remain so when they are referred to any other system which, in comparison with the first, is executing a uniform rectilinear motion. The coordinate systems in which the law of inertia applies are referred to as inertial systems.

One must be careful to distinguish quite clearly between these inertial systems, and the many other systems for which the law of inertia is not valid. Such systems are all those which have an accelerated, or a rotatory motion with respect to the first. Strictly speaking, this applies to the motion of the earth. In fact we see the Trade Winds departing from their rectilinear path. Let us return to the same experiment which we have just considered and assume that during the time when the sphere is falling, the ship, urged on by some tremendous wave, makes a sudden jerk. Then the sphere will no longer fall in a straight line but will perhaps strike against the mast if conditions are suitable. But if on some other day this ship is at rest and the observer's receives a sudden jerk, will the sphere strike against the mast now? It will not. From this different behaviour, *i.e.* from the hitting of the mast when its own ship suddenly moves, and the absence of such when the other ship is jerked forward, one can recognize with certainty whether the one or the other ship moved suddenly. If, therefore, the sphere fell according to the law of inertia in a straight line, the motion would no longer be rectilinear when observed from the ship which was suddenly jerked. This latter ship would therefore no longer be an inertial system, and the law of inertia would no longer be valid for this system of coordinates.

3. Momentum and Impulse

When a force produces acceleration in a body for a considerable time, it is not difficult to determine the velocity and acceleration, from observations made on the path traversed. But it may happen that the force only acts for such a short time, that, during this interval, no observations concerning the path described can be made. The primary cause of our inability to do this lies in the slowness with which our senses comprehend the velocity of natural events which are being considered. If, for example, a billiard ball is struck with a cue, the acceleration of the ball during the blow cannot be followed; we only observe the total effect of the blow, *i.e.* the appreciable change which has taken place in the velocity of the ball. Let us therefore investigate the connexion between the force and the total change in the velocity.

We will assume, for the moment, a constant force which acts for t seconds upon a freely moving object. The equation $f = ma$ is valid. Every second the velocity changes by the same amount; consequently, after t seconds, the velocity is changed by an amount at. Multiplying the above equation throughout by t, we have

$$ft = mat$$

Let v_0 be the initial velocity; then after t seconds it is given by

$$v_1 = v_0 + at, \text{ or } at = v_1 - v_0$$

We may therefore write

$$ft = m\,(v_1 - v_0) \qquad (9)$$

If the initial velocity were zero ($v_0 = 0$), the relationship between the force, f, and the velocity acquired after any time t, would be

$$ft = mv \qquad (10)$$

From the mass and the change in velocity, no conclusion can be made concerning the magnitude of the force, but only concerning the product of the force and the time for which it was applied. This product is called the *impulse*, whilst the product mv is termed the *momentum*. We have therefore proved, for this particular example, that the momentum is equal to the impulse.

Now let us consider a very transient impact. So quickly does such an impact take place, that we have no right to assume that the force has been constant during the short time for which it lasts. But if this very brief interval of time is further subdivided into a very large number of short time elements, these can be made so short, that during any one of them the force does not change appreciably. We may therefore consider the force, during every one of these small time intervals, as constant, and apply the above formula (9). If we denote the different small periods of time by $t_1, t_2, t_3 \ldots$; the corresponding forces by $f_1, f_2, f_3, \ldots$ and the velocities at the ends of these times by $v_1, v_2, v_3, \ldots$ we have the following equations:

$$\begin{aligned} f_1 t_1 &= m\,(v_1 - v_0) \\ f_2 t_2 &= m\,(v_2 - v_1) \\ &\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ &\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ &\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ f_n t_n &= m\,(v_n - v_{n-1}) \end{aligned}$$

On the right-hand side of these equations appear terms which are multiples of the increases in velocity, which occur consequently, so that the only logical process is to add them together. On the left-hand side stand the small elements of impulse. By adding all these equations together we obtain:

$$f_1 t_1 + f_2 t_2 + f_3 t_3 + \ldots + f_n t_n = \Sigma\, ft = m\,(v_n - v_0). \quad . \quad (11)$$

an equation from which the velocities $v_1, v_2 \ldots v_{n-1}$, have vanished completely.

If the sum of all the small elements of impulse is again called the impulse, the law for variable forces can therefore be stated: **the change in momentum is equal to the impulse.**

It is easily seen that this law of impulse will be of great importance in problems concerning impact, and similar events.

4. Action and Reaction

Our ideas concerning force are substantially strengthened by the fundamental law of action and reaction which was postulated by **Newton.** It was first elucidated by Newton himself, with the aid of the following example. If the tip of my finger is pressed against the top of a table, my finger is thereby flattened; an occurrence which can only take place if something presses against my finger. In the above example it can only be the table, for this is the sole object in contact with my finger, at that instant.

If a horse drags a heavy stone forward by means of a rope, this latter is drawn taut, and slightly elongated. Its tendency to contract again shows itself at the end near the stone in a force, which draws the stone forward, and at the other end in a force tending to hold back the horse.

These relationships can be made more clear if, in place of the rope, one considers a rubber tube or spiral spring. Such a spring, Fig. 10, is stretched **when it is pulled at both ends in contrary directions.** If one desires to move a stone forwards by means of such a spring, the strong tendency for the spring to elongate, is a sign that the two ends are being drawn by forces tending to separate them.

Such a reaction will also be exerted on a body which is free to move if it is acted upon by a force tending to accelerate it. If we endeavour to set the carriage of our suspension railway in motion quickly, by means of a light spiral spring attached at one end to the carriage, whilst a force is applied at the other, it will be observed that the spring undergoes a marked extension, a proof that its two ends are being pulled in opposite directions.

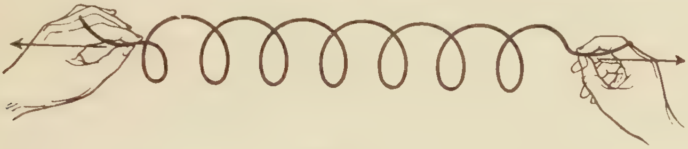


FIG. 10.

The following example, likewise chosen by Newton, also enables us to obtain a very instructive mental picture of the processes which are at work. From two strings there are hung two spheres which, when at rest, just touch (Fig. 11). If the sphere a is drawn out to the position a_1, and then released, it will strike the sphere b which was formerly at rest. After the collision it will be observed that the sphere b has been suddenly set in motion, and moved outwards to the left. But since it is accelerated towards the left it will exert a force upon a which is directed towards the right, in virtue of which the sphere a is accelerated in the opposite direction. The effect of this reaction upon a is to cause its motion towards the left to be checked considerably. It depends upon the ratio of the masses a and b whether, after the impact, the sphere a moves towards the left with diminished velocity, or even flies backwards. The experiment is of especial significance if two equal steel spheres are utilized, for the sphere a is just brought to rest after the impact. But we know that a force is necessary to bring a moving body to rest and, in the present case, this can only originate through the sphere b, which is the only body in contact with a.

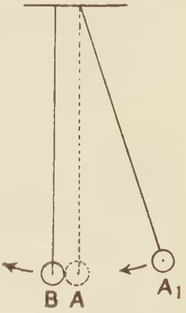


FIG. 11.

Now at every moment during the impact a force, equal to that tending to move the sphere b to the left, tends to move a to the right, each body therefore experiences an impulse, the corresponding changes in momentum, according to the foregoing remarks, being equal in magnitude, but opposite in sign. If m is the mass of a, m_1 the mass of b, whilst u and u_1 are the velocities before, and v and v_1 the velocities after, impact, then

$$m\,(v - u) = -\,m_1\,(v_1 - u_1) \quad . \quad . \quad . \quad . \quad (12)$$

From this formula it is at once apparent, that the change in velocity is inversely proportional to the two masses.

Let us now try to understand this fundamental law of mechanics a little more fully. Our first ideas concerning forces were certainly acquired from experience associated with our own bodies. We perceive the exertion of our arm when throwing a stone, from the sense of feeling, and we notice the pressure upon our hand which is thereby exerted by the stone. We also know very well that we can determine the instant, direction and velocity of the stone, which has been thrown. It is we who first act upon the stone, thereby evoking the reaction. We are very much inclined to apply these ideas concerning force to inanimate objects without any further consideration. In the above example, for instance, we see the one sphere hanging at rest and the other drawn aside from its position of rest, and then released; we observe it moving towards the sphere at rest until the collision occurs, and are inclined to designate the effect of *A* on *B* as the action, whilst the effect of *B* on *A* is called the reaction. This, however, is entirely an arbitrary assumption. We know that our own earth rotates about its axis, and that it moves around the sun at the same time once every year. Might it not therefore happen that, at some one instant, these motions were exactly opposite and equal to that of the sphere *A*? Then the sphere *A*, at the instant when the impact occurred, would be just at rest, and *B* would apparently be hurrying towards *A*. By raising the sphere *A* we should have brought it to rest so that *B* must have made impact on the stationary sphere *A*.

Hence, what we have previously called reaction, would become the action, and conversely. Since we can never know this, we can never differentiate between action and reaction. It would therefore be preferable to use an expression which does not imply that one sphere acts before the other and merely say that the **two bodies have collided.**

If a heavy object lies upon the table, it acts upon the latter in virtue of its weight: that is the action; it presses the fibres of the wood together a little, so that they tend to expand again and exert a pressure upwards; that is one conception of reaction. In reality the weight of the body is due to the mutual attraction between the body and the earth. From what has already been stated, we know that the earth will be attracted just as much from above, as the objects upon the earth's surface are from below. The whole phenomenon therefore consists of the object and earth endeavouring to approach one another more closely, the intervening table preventing this from taking place. The force which acts from above is, in reality, participating in the phenomenon just as much as the one from below; in fact we are not justified in designating the one as the primary force in preference to the other.

This mutual action between two bodies which are endowed with the same physical property is, it would appear, an essential characteristic of all forces in the inanimate world. An inert body can only attract another inert body. A moving body can only act upon another moving body on account of its own inertia and that of the second body. An electrically charged body appears, however, to attract every body, including those which are not charged, but careful investigation has shown that, in this case, the uncharged body is first charged electrically by induction, and is then attracted. A magnet appears to attract a piece of unmagnetized iron, but the moment the attraction begins the iron becomes a magnet, and substances which can only be magnetized a little, or not at all, will therefore experience a little, or no attraction whatsoever. This fundamental law is without exception true, so that we are able to conclude that there exists, in each of two participating bodies, that particular property with which we are dealing, solely from the fact that a mutual action occurs between them. When, therefore, **Zeeman** had demonstrated the influence of magnetic and **Stark** that of electric fields upon the emission of light waves, all physicists were thereby justified in assuming that light waves had been proved to be essentially electromagnetic in character.

That the two actions are absolutely equal to one another was proved by Newton himself, by transferring both actions completely to a third object. He was then able to show that they always compensated one another exactly. For example, let a magnet be placed upon a piece of cork, and at a little distance away, but upon the same piece of cork, a piece of iron, so that the iron tends to produce motion in one direction just as much as the magnet does in the other. If the whole is placed on water, a very small excess of one force over the other would easily become visible, for the whole would move in the direction of the greater force, but in this case not the slightest acceleration can be detected.

5. WORK AND ENERGY

The physical idea of work is formed in connexion with the same idea in daily life. In fact, the word can also be used for mental work, but we do not have to consider that aspect now. When it is a question concerning work on material bodies, the application of a force is always included, and the work done is greater, the greater the applied force becomes.

The application of a force can be referred to in two ways ; in the first, equilibrium is maintained between a force which is already present, and the one which is applied, whereas in the second a body moves notwithstanding the existence of an opposing force. The first occurs, for example, when a mass must be supported so that it does not follow gravity and fall. One is soon convinced that it is easier to do this, if a table or trestle is placed underneath the body. That will be effective for as long as is desired, and does not become tired. Nobody would say that the table has done any work, no more than they would say that a stone in the foundation of a building does work, because it supports the brickwork which lies above.

But it is different when a body **moves** against a force. The bricklayer, who carries the stones when a building is under construction, does work, and the amount of this work cannot be lessened in any way by the use of some contrivance which cannot move. If the stones are carried up five metres, the labourer does a certain amount of work for every stone that he carries. If, later on, one of these stones is carried ten metres, the work can be divided into two amounts each equal to the work done in carrying the stone five metres, and the work will be all the greater, the greater the distance through which the force of gravity has to be overcome.

At first some quite arbitrary amount of work is chosen as the unit, so that the amount of work, W, which is performed where a force of 1 *dyne* is overcome through a distance 1 *cm.* is, let us say, C units ; the amount of work necessary in order to overcome a force of F *dynes* through a distance s *cm.* is therefore

$$W = CFs \quad . \quad . \quad . \quad . \quad . \quad (13)$$

The magnitude of C depends only upon that of the unit of work which has been chosen. If this unit is so chosen that unit work is done when a force of 1 *dyne* is overcome along a distance of 1 *cm.*, then is $C = 1$, and we have for the work done when a force F is overcome along a path of s *cm.*

$$W = Fs \quad . \quad . \quad . \quad . \quad . \quad (14)$$

This unit of work is the *erg*, and has dimensions [*c.f.* equation (7)]

$$W = [m\, l^2\, t^{-2}] \quad . \quad . \quad . \quad . \quad . \quad (15)$$

If the direction of motion does not coincide with that in which the force is applied, as with an inclined plane, the force is resolved into two components, one in the direction of the motion, and the other perpendicular to it. If the force and path include an angle d, so is F', that component of the force acting in the direction of the motion, given by $F' = F \cos a$ (Fig. 12) and consequently the work done is

$$W = F' l = F l \cos a = F h, \quad . \quad . \quad . \quad (16)$$

where l and h are the length and height of the plane respectively.

A certain expenditure of energy is also necessary if a body is to be set in motion, even if it is not held fast in position but is quite free to move; for we have seen that, on account of its inertia, a body can only be caused to move by the application of a force. Since this force is always operative for a certain time, and during this time a certain distance is covered, the acceleration of a body must therefore be inseparably connected with a performance of work. The amount of work necessary is very easily obtained. If the force F moves through a distance s the work done is

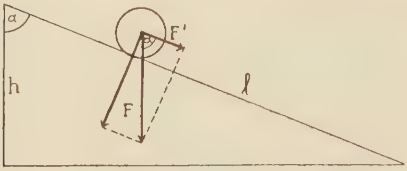


FIG. 12.

$$W = Fs$$

If the force is constant the body receives a uniform acceleration, a, given by $F = ma$.

Since the motion is one of uniform acceleration the path traversed is given by $s = \frac{1}{2} at^2$, from which it follows that the work done is expressed by

$$W = Fs = ma. \frac{at^2}{2} = \tfrac{1}{2} mv^2 \quad . \quad . \quad . \quad . \quad (17)$$

where v denotes the final velocity after time t.

This result is very remarkable, for the acceleration a and the time t for which the force is applied are not individually of prime importance,—they do not occur explicitly in the formula, which expresses the work done; this depends rather upon the velocity which is finally imparted to the body. The equation shows, therefore, that if a body of mass m is to acquire a velocity v, then the amount of work necessary for this purpose is $\frac{1}{2} mv^2$. Since the product of the force and the distance traversed must always have this value, it follows that the smaller the force, the greater must be the path traversed before a body can acquire a definite velocity.

In the above we have distinguished two effects due to a force. In the first case the force may hold another in equilibrium; in other words a force can prevent a body from being accelerated by another force. The second effect occurs when the body, in spite of the opposing force, moves. If these two effects are examined more closely, the following difference is detected: if a force holds another in equilibrium, *i.e.* without producing motion, no trace of this effect is left behind. However long the equilibrium may be maintained, everything is afterwards just as it was previously. This is not so if the force has been applied through a certain distance. For on some occasions the body may have been raised to a higher position, whilst on others it may have a greater velocity. This is a very important point.

Energy. If a body is situated a certain distance above the earth's surface it has, in virtue of its position, something which the same body did not possess before it had been raised. If it is connected to another weight, by means of a rope passing over a pulley, Fig. 13, this latter weight can be raised; in the process, however, the first body forfeits its higher position. As a further example we may cite the water which, falling from a height, is used to work a miller's wheel and drive the mill. It may also fall on its own accord, and acquire a certain velocity; in short, it can do that which we have learnt also to recognize as **work.**

Similar things may occur when a body is in motion. If it strikes against another body it imparts to this latter a certain velocity. If the carriage (Fig. 8, p. 25) is made to move towards the right with a certain velocity the weight W is raised, and a certain amount of work done. If an object is thrown upwards, the force which must be overcome is the weight of the body itself. If will be noticed that wherever work is done by a body its own velocity becomes less. At the instant when this velocity is reduced to zero the body is no longer able to do work. We find, therefore, that at least in those examples which have been considered, a certain amount of work must be performed upon the body before it can again do work itself. The capacity of a body for doing work is called **energy.** It is measured by the amount of work the body can do and is expressed in the same units, and has the same dimensions, as work itself.

Let us now examine these relationships more closely. If a body is attracted by the earth with a force F and, in spite of this force, it is raised a distance h, the amount of work done by the force is Fh. This work is performed, for example, in raising a quantity of water of corresponding mass to a container placed at a height of h cm. The hydrostatic pressure of the water upon the bottom of the container has not been changed thereby, but the water can now return to its former position and do work even whilst it exerts this same hydrostatic pressure upon the vessel in which it is stored. The capacity which a body has for doing work is intimately connected with the force tending to cause the body to fall. But this force is still present when the mass has fallen, and is no longer able to do work. Now just as a force is essential, when work is to be done, it is also necessary for the point of application of this force to move through a certain distance, as, for example, when a weight is raised to a more elevated position.

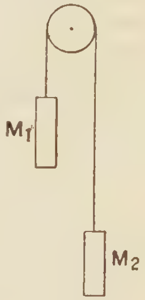


Fig. 13

We have therefore established the fact that a certain amount of work has been performed on a body, when this latter has been raised above the earth's surface. The work necessary for this is Fh. In consequence of this higher position the body is now capable of doing work the magnitude of which is Fh, a quantity which is exactly equal to the work which must be expended in raising the body to this position. The body can remain in this position as long as is desired, but its capacity for doing work is not thereby affected. From this fact one concludes that the work necessary to raise an object is not spent, but still exists in the raised body. This capacity for doing work, which depends upon the position of the body, is called the **energy of position** (or **potential energy**).

It is now easily seen, that all these statements are true for forces which endeavour to change the position of a body, whether they be electric, magnetic, or elastic, in their origin. In order to separate an electrically charged body from one charged with electricity of opposite sign, to withdraw a piece of iron from a magnet, or to raise a stone above the earth, work must be expended. When such pairs of similar bodies have been removed to a considerable distance from one another they become capable of doing work, when they return to their former position. There are therefore just as many kinds of energy due to position, as there are different kinds of forces, with which bodies can exert an influence upon one another.

The capacity for doing work, which a body possesses in virtue of its velocity, is somewhat distinct from all these other types of energy. It is called the **energy of motion**, or **kinetic energy**. Let us apply our ideas again to the suspension railway (Fig. 8, p. 25). We will assume that the carriage has a mass m and that it acquires a velocity v through the application of some force, the direction of which is such that the carriage moves to the right and thereby raises the weight W. The force necessary for this is W, the weight of the small mass. This force retards the motion of the carriage, the retardation being expressed by $a = \frac{F}{m} = \frac{W}{m}$. Every second the velocity of the carriage decreases by an amount a, so that after t secs. this will be at, and when $at = v$ the carriage will have come to rest. This occurs after a time interval $t = \frac{v}{a}$ secs. We can now calculate the path traversed by the carriage before coming to rest. Since the motion has been retarded uniformly, the path traversed in t secs., which is also the highest through which W has been raised, is

$$h = \tfrac{1}{2}\, at^2 = \frac{v^2}{2a}$$

The amount of work performed is

$$Wh = F\frac{v^2}{2a} = \tfrac{1}{2}\, mv^2$$

A very important result is obtained from this equation : in consequence of its velocity v, a body, of mass m, can do work, equal in amount to $\frac{1}{2}\, mv^2$ when the body has completely lost its velocity. But this is exactly that amount of work which must be expended upon the body, in order for it to acquire this velocity.

We therefore arrive at the very important result that the work expended in causing an object to be accelerated is not spent, but is still available in the state of motion of the body.

From a few mechanical examples which we here survey we shall comprehend the meaning of the conservation of energy. The conservation of energy, for purely mechanical processes, is a necessary consequence of the laws of force and inertia. No new experiments are necessary to establish this, for it can be deduced from the laws which have already been proved.

Hitherto we have always assumed the force to be invariable along the whole distance through which it is applied. But that is not necessarily so. A variable force is capable of doing work. We may assume, that, over the small distance Δs, the force F remains constant so that the work done is $\Delta W = F\Delta s$. When the force becomes F_1 which is constant over the corresponding distance Δs_1, the work done is $\Delta W_1 = F_1\,\Delta s_1$, and if such changes occur frequently, the total work done is $W = \Sigma F\Delta s$, where the expression is summed up over all the distance traversed. If the force changes continuously we have, in the language of the calculus, $dW = Fds$, the total work being $W = \int Fds$.

In the simplest case the force is proportional to the distance, *i.e.* $F = Cs$, the force being zero in the beginning, and increasing regularly with the distance. For each centimetre traversed the increase in the magnitude of the force is C, and at the end of the distance, s_1, it has become $F_1 = Cs_1$, so that

$$W = \int_0^{s_1} C\,s\,ds = \frac{Cs_1^2}{2} \quad . \quad . \quad . \quad (18)$$

The work done is therefore just the same as if, instead of a variable force, acting throughout the whole distance, it had been equal to half the final value. For this would be $\frac{F_1}{2} = \frac{Cs_1}{2}$, and the work done, when the point of application of this force moves a distance s_1, is

$$W = \frac{F_1 s_1}{2} = \frac{Cs_1^2}{2}.$$

6. MOTION IN A CIRCLE

The laws which have already been discussed can be applied to all types of motion ; there is no motion which is not governed by these laws. We shall have to deny ourselves a further discussion of this almost infinite variety of possible motions, and will only discuss two examples, which will be used repeatedly in the chapters which follow. We refer to circular and vibratory motions.

Circular motion had already been referred to by the ancient Greek philosophers. They had not failed to notice that the heavenly bodies moved in circular orbits. When they observed the majestic tranquillity of these movements, and saw how stedfastly the heavenly bodies completed their paths, returning again and again to the same point in space after every complete revolution, it appeared to them that circular motion was the most perfect of all motions, because, as one may say, it combined tranquillity with movement. Since these people considered the eternally shining firmament as belonging to a higher order of things, it appeared to them to be quite expedient for such a higher order to be endowed with the most perfect of all motions. To consider any further possible laws of circular motion did not occur to them. It was only when it had been recognized that the laws of dynamics are quite general, and the differences between natural and forced motions had been settled, that the special laws of circular motion began to be discussed. The Dutchman, **Christian Huygens,** appears to have solved this problem at the same time as **Newton** did in England.

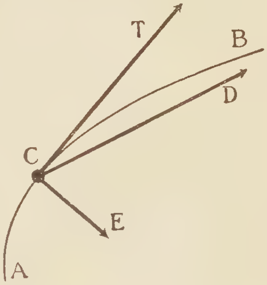


FIG. 14.

Let us consider first of all a body, Fig. 14, which is moving along a curved path AB, according to no particular law. Since the path is nowhere rectilinear, a force must act continually upon the body, and prevent it moving along a straight line tangential to the curve. Let CD represent the acceleration at any point C. This acceleration can always be resolved into two components, one being directed along the tangent T, whilst the other, CE, is perpendicular to it. The component of the acceleration along the tangent will increase the tangential component of the velocity ; the other will cause the body to deviate from its otherwise rectilinear path. In our present problem, where the motion has to occur along a circular orbit with constant speed, this first component must be zero at all times. The force must therefore be perpendicular to the orbit everywhere, *i.e.* it is directed towards the centre of the circle. It does not increase the speed of the body, but only keeps the body upon its correct path, by providing the necessary curvature to the orbit. Since the curvature of a circle is everywhere constant, it follows that, for a uniform motion in a circle, a force must be exerted upon the body ; this force is constant in magnitude and directed towards the centre of the circle. If we consider several bodies all revolving in the same circle, with the same velocity, they must all experience the same deviation in the same time whatever their mass may be. The forces which are necessary will, however, be proportional to these masses. We have therefore only to discover the centripetal acceleration which is common to all these bodies. Let the body be at some point, A (Fig. 15). Without the application of any other force it would proceed in the direction of the tangent with velocity v, and in a small time t it would arrive at B, where $AB = vt$. But through the continuous action of the constant force it never gets to B, but comes to some point below B. The greater the force directing the body towards O is the further from B will it move. We must now inquire how great this force must be so that the point C shall lie exactly upon the circle. If $AB = CD = h$ and $BC = AD = e$, the following equation is true,

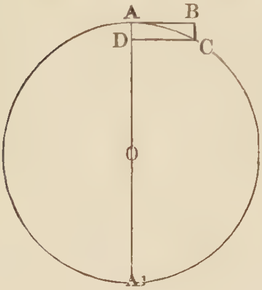


Fig. 15.

$$\frac{e}{h} = \frac{h}{2r - e}$$

or, more approximately, if e is small compared with $2r$,

$$e = \frac{h^2}{2r} \quad . \quad . \quad . \quad . \quad . \quad (19)$$

In this equation e is the distance, which the body traverses in time t, under the action of a constant force directed towards O. If a is the required acceleration, then this distance is given by $e = \frac{1}{2}at^2$. Now the distance AB is $h = vt$, whence it follows from equation (19) that

$$\frac{at^2}{2} = \frac{v^2t^2}{2r} \text{ or } a = \frac{v^2}{r}.$$

Now the velocity of the rotating body depends upon T, the time of a complete revolution, and the relationship is expressed by the equation $v = \frac{2\pi r}{T}$, so that one can therefore write $a = \frac{4\pi^2 r}{T^2}$.

If a is known this equation enables T, the periodic time, to be calculated. The result obtained is

$$T = 2\pi \sqrt{\frac{r}{a}} \quad . \quad . \quad . \quad . \quad (20)$$

Instead of the periodic time a new idea may be introduced, viz. the angular velocity. In the time T a point which is distant 1 cm. from the axis of rotation, traverses a distance 2π, so that the path described in one second is $\frac{2\pi}{T} = \omega$ (say). This is called the angular velocity of the rotating body, so that we may write $a = r\omega^2$.

Finally, especially for very rapidly rotating bodies, it is sometimes preferable to replace the periodic time by the number of revolutions made in 1 second. If T is the periodic time, the number of revolutions per second is $\nu = \frac{1}{T}$, so that $v = 2\pi r\nu$.

If we collect together these various expressions for the centripetal acceleration, we have

$$a = \frac{v^2}{r} = \frac{4\pi^2}{T^2}r = r\omega^2 = 4\pi^2\nu^2 r \quad . \quad . \quad . \quad (21)$$

Centrifugal Force. As always happens, whenever a body is accelerated, a reaction is also produced, which is equal in magnitude to the accelerating force, but which acts in the opposite direction. In the present example its action is directed away from the centre of the circle, so that it is called the centrifugal force. There can be no possible doubt about its existence, for if a stone attached to one end of an elastic band is caused to swing in a circle, the elongation of the band is clearly seen, and this is an indication that it is being pulled at both ends by forces, which act in opposite directions. If the sling is not sufficiently strong it breaks. Similarly a rapidly rotating wheel may fly to pieces. The centrifugal pump, the milk separator and other applications of everyday life depend upon the utilization of centrifugal force. Nevertheless this idea is often the subject for dispute, and it is to be admitted that, in the more difficult processes, it is not always easy to say how the centrifugal force manifests itself.

For the present we must consider what was said above concerning action and reaction ; strictly speaking, we cannot distinguish between which is to be called the action, and which the reaction. If, for example, a glass flask, half filled with water, is made to rotate very rapidly, the water presses against the sides of the flask, rises up in the same, and finally forms a belt or girdle which appears to hover in the flask in spite of gravity. The water endeavours to move outwards in consequence of its inertia. But it is constantly deflected by the walls of the vessel towards the centre of its circular path ; the centrifugal force shows that the water presses against the glass walls, so that if the flask is not sufficiently strong it would break. But it is now obvious that this pressure of the water, in consequence of its inertia and the curvature of the glass walls, is of prime importance, and the deflection towards the centre is only the reaction produced by the glass against the pressure of the liquid. We cannot say which of these actions is produced first, for they are always simultaneous ; the pressure between the water and the walls of the flask, in so far as it is exerted upon the walls, is the centrifugal force, but in so far as it is exerted upon the liquid, it is a centripetal force. Both actions therefore always cease simultaneously, *i.e.* at the instant when the object ceases to move along a circular path.

When **H. Hertz** raises the difficulty that at one time we call inertia what later on we call centrifugal force, we must reply that by the term inertia we understand the capacity of a body for acting like a force upon another moving body. If it is desired suddenly to set in motion a body which is free to move, such as a 5 kgrm. mass suspended by a long cord, and a spiral spring or indiarubber tube is used for that purpose, the spring is considerably extended because the mass has inertia. And yet the extension of the spring indicates the action of a force on the part of the freely hanging weight, which is nevertheless accelerated.

7. RECTILINEAR (NORMAL) HARMONIC MOTION

Let us assume that a mass M, Fig. 16, is held in a definite position by some forces (let us say two springs). If the mass be displaced from its position of rest, a force is called into play tending to draw the body back to this position, and this force becomes all the more noticeable, the greater the displacement from the position of rest. If the mass M, whilst at rest, receives a blow which imparts to it a velocity v, it will be pulled back with a greater force the further it is displaced from its zero position. It will therefore soon come to rest, let us say after a displacement OA, and immediately seeks to return to its position of rest, which it reaches with an ever-increasing velocity. In consequence of its inertia, however, it passes through this position to the other side, and is thereby restrained by a force, which increases with the displacement, until, finally, it comes to rest again at B, let us say. The particle then returns, passes through the zero position, and so the motion proceeds—a continuous alternation between the force and inertia of the whole system.

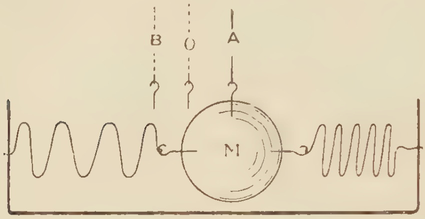


FIG. 16.

In Nature there is a great number of such bodies, which are attracted to some zero position by a force which is proportional to the distance, and the motions of such systems form the basis for many phenomena.

In order to describe this motion more accurately the above type of motion is coupled with one taking place uniformly in a circle. For that purpose we use the law concerning the independence of motions (p. 27). We can express this law as follows: A particle moves with any arbitrary motion upon a curved path, the shape of which is quite optional. Fig. 17. At the same time a second particle moves along the straight line CD, so that when the sun shines vertically upon the first particle, the second is always in its shadow; expressed somewhat differently, one may say that if a vertical straight line is drawn from the first particle upon CD, then the second particle is always to be found at that point, where this vertical line meets CD. If, therefore, the motion of the first particle is described by means of coordinates x and y, and that of the second by x' and y', then must $x = x'$ and $y' = 0$ at all times. Thus the motion of the second particle is in no way whatsoever dependent upon how far the first one moves upwards or downwards, but only how far it moves towards the right or left. This fact enables us to make two assertions: firstly, $x' = x$; and secondly, if v and v' are the velocities of the two particles, the velocity of the former may be resolved into two components v_x and v_y, so that v' must equal v_x. If this were not so the second particle would move a greater or less distance towards the right than the first one does, an occurrence which would violate the conditions of our problem. Thirdly, from a similar reasoning, it follows that the acceleration of the particle moving along CD at any instant

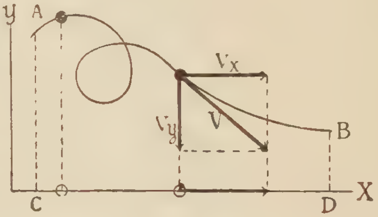
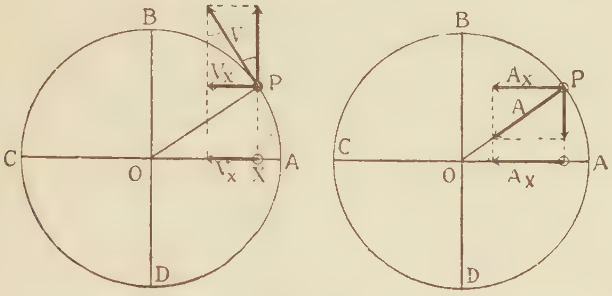


FIG. 17.

must be equal to the x-component of the acceleration with which the second moves along the curved path, for otherwise, in the next instant, the velocity would no longer have the appropriate value, and therefore the position upon the straight line would no longer correspond with the assumptions we have made.

The above process is now applied to the case when the curvilinear motion is a circular one. Let the origin of coordinates be the centre of the circle. The rotating particle moves with uniform speed along $ABCD$, Fig. 18, whereas the rectilinear motion is performed by a particle moving to and fro along AOC. When the first particle is at A, the second is there also. If the first is at B, the second is at O; afterwards they are both together again at C, and when the first particle arrives at D, the second is back at O, and then travels along OA, while the first moves along the quadrant DA to the starting point.



(a) Fig. 18. (b)

The position of the revolving body, at any time t, can be ascertained by joining P to the centre of the circle with a straight line, which includes with the initial line OA an angle which increases regularly with the time. If ω is the angle described per second, at any time t, it becomes ωt, and the x-coordinate of the point at any time t is $x = r \cos \omega t$. This must also be the position of the point upon the straight line, *i.e.*

$$x' = r \cos \omega t \quad . \quad . \quad . \quad . \quad (22)$$

The velocity, v, of the rotating particle is not changed in magnitude, Fig. 18 (a), but it changes its direction continuously (it is always along the tangent at the point in question), so that its x-component is a variable, it being in fact

$$v_x = v \sin \omega t \quad . \quad . \quad . \quad . \quad . \quad (23)$$

The velocity of the particle moving upon the straight line has this same magnitude at any instant. Both bodies are at A when t is zero, so that v_x is zero at that point; at B, $\omega t = \pi/2$, so that v_x attains its maximum value there. The velocity of the particle moving along the straight line AOC varies between the limits $+ v$ and $- v$.

The acceleration, a, of the rotating particle [*cf.* Fig. 18 (b)], is likewise constant in magnitude but not in direction. The x-component is

$$A_x = - a \cos \omega t = - \frac{a}{r} x \quad . \quad . \quad . \quad . \quad (24)$$

This equation signifies that the particle which moves upon the straight line must have an acceleration which is directed towards the centre and the magnitude of the acceleration must be proportional to the displacement of the object from its position of rest. Since we had assumed the particle to be one which is continuously accelerated in this manner, the problem of finding its position and acceleration has been solved. The equations (22), (23) and (24) specify the position, velocity, and acceleration at any time t, of a particle which moves according to this law. This type of motion is called a **rectilinear or normal harmonic motion.**

The greatest distance of the oscillating point from its zero position is called the *amplitude*; if it is denoted by A, then r must be put equal to A in the above formulae.

The position of a body vibrating in this manner can therefore be found at any time t, by describing, with centre O, a circle of radius A, and imagining that a subsidiary body moves regularly around this circle with angular velocity ω. If the maximum value of the velocity is denoted by v_0 whilst v is used to denote the velocity at any other time, and a_0 is likewise the maximum acceleration, the equations which give the motion at any time t are as follows:

$$\left.\begin{array}{ll}\text{The distance from the position of rest,} & \\ & x = A\cos\omega t = A\cos\frac{2\pi}{T}t, \\ \text{the velocity} & v = -v_0\sin\omega t = -v_0\sin\frac{2\pi}{T}t, \\ \text{the acceleration} & a = -a_0\cos\omega t = -a_0\cos\frac{2\pi}{T}t.\end{array}\right\} \quad (25)$$

From these formulae it is seen at once that the position and velocity of an oscillating point are determined, in the first place, by the angle ωt; so often as the angle increases by an amount 2π, the vibrating point has moved through all positions from $+a$ to $-a$ and back again. The angle ωt is called the **phase angle**, and the position of the body, in so far as it is defined by this angle, is called the **phase** of the body. The interval of time in which a complete to and fro motion takes place is called the **period of oscillation.** It is obviously equal to the period of the rotating body, which has already been given by equation (20), viz.

$$T = 2\pi\sqrt{\frac{r}{a}}$$

In this equation r must be replaced by the amplitude A, and the acceleration by a_0 which is the maximum value of the acceleration. We find therefore

$$T = 2\pi\sqrt{\frac{A}{a_0}} \quad . \quad . \quad . \quad . \quad (26)$$

But since, for different values of A, the corresponding acceleration changes in a similar ratio, $\frac{A}{a_0}$ is independent of the displacement. Hence it follows that the periodic time always has the same value. Accordingly, the periodic time of a system executing simple harmonic motion can be stated, as soon as the acceleration at any point removed from the zero position is known. If, for example, the acceleration at a distance of 1 cm. is called a_1, the periodic time is therefore

$$T = 2\pi\sqrt{\frac{1}{a_1}} \quad . \quad . \quad . \quad . \quad (27)$$

The value a_1 alone, is sufficient to determine the periodic time. If the force, F, is given, whereby the body is pulled backwards towards its zero position, the acceleration is related to the force and mass m according to the law $F = ma$.

Two characteristics must therefore be possessed by a system if it is to be able to execute simple harmonic motion ; a force must exist which is always directed towards the same point, and which is greater the greater the displacement from the zero position ; there must also be a mass having inertia. The force always accelerates the moveable body towards the zero position ; the inertia carries it beyond this position, thereby calling the force into play again.

Considered from the point of view of energy, there is a continuous exchange of potential energy to kinetic energy, and *vice versa*. The former is due to the distribution of the forces in the system, the latter to its inertia. If a loss of energy is avoided, the inertia is to be considered as the only resistance which has to be overcome by the forces in question. For one can see then that the most extreme position is again reached by the moving body after a complete swing. The kinetic energy, which the particle possesses at any position, is equal in magnitude to the potential energy which has been lost. If, then, ma_1 is the force tending to cause the body to return when it is displaced 1 cm., and if A is the displacement, then at distance A the force is $F_0 = ma_1A$. Since the path traversed is A cm., the loss in potential energy is $W = \frac{1}{2}ma_1A^2$; the kinetic energy is greatest when the body passes through its position of rest and amounts to $\frac{1}{2}mv_0^2$. Since these two expressions are equal, it follows that

$$v_0^2 = a_1A^2 \text{ or } v_0 = A\sqrt{a_1} \quad . \quad . \quad . \quad . \quad (28)$$

This equation signifies that the velocity which is a maximum when the displacement is zero, and the displacement which is a maximum when the velocity is zero, bear a fixed ratio to one another ; this ratio is $\sqrt{a_1}$.

In actual vibratory motions the problem is not yet completely solved by the consideration of the interplay between force and inertia. Some other disturbing factor is always present. Very frequently this can be reduced to negligible dimensions, so that one is justified in neglecting it, but it never quite vanishes, and frequently it becomes so considerable that it changes the whole phenomena. We refer to the viscous resistance which the moving body encounters in virtue of its motion ; sometimes it may be due to the air which has to be pushed aside ; often it is due to axial friction in the apparatus ; or again it may arise from the sagging of a wire, cord, etc. By all these influences the motion of the body is always impeded ; it is never increased. Consequently, the vibrating body is often brought to rest somewhat earlier than the above considerations would indicate. Each deflection becomes less than the preceding one, and the whole motion is brought more or less quickly to rest according to the magnitude of this force. Such a motion is termed a **damped oscillatory motion**. The whole course of such a motion can be represented by means of a formula, which differs from the one above only by the fact that the amplitude becomes smaller as the time increases. It is

$$x = A\,e^{-ct} \cos \omega t \quad . \quad . \quad . \quad . \quad . \quad (29)$$

where e is the base of natural logarithms ($e = 2{\cdot}7183 \ldots$), and the factor c is due to the frictional forces.

The friction may even be so large that the body does not execute one complete swing, but returns straight to rest after having been displaced a distance A.

CHAPTER III

GRAVITY AND GRAVITATIONAL FIELDS OF FORCE

1. Gravity

THE weight of a body has often been referred to as a force, and it is now necessary for us to examine this statement a little more closely. The facts that bodies left to themselves tend to fall downwards, and that, when held in the hand, they exert a pressure upon it, are facts which could not have remained latent to mankind for any length of time. Yet thousands of years elapsed before man was able to obtain the correct view concerning these simple phenomena of everyday occurrence.

We have already seen how, in the Middle Ages, following a precedent of the Ancient Greeks, some bodies were imagined to have a tendency to fall to the earth, whereas others tended to leave it. This tendency existed only on the part of the moving body, just as, for example, the desire to eat only occurs in a person who is hungry. Man has always inclined straightway to apply the experiences of his inner life as a model for his understanding of the outside world. But since it has become recognized that bodies do not move by virtue of some internal desire, but only as a result of an external agency, it has gradually become customary for the weight of bodies to be conceived as arising from some external agency tending to cause motion. The direction of this force, which is clearly directed towards the centre of the earth, indicated that the earth was the origin of this force.

Until recently the view was generally held that the fundamental laws of motion, and especially those of falling bodies, were discovered by **Galileo.** More exhaustive inquiries, for which **Duhem** deserves special thanks, have irrefutably proved that, in the main, Galileo only extended what others had already begun. Another conception concerning moving forces, which was in very definite contrast to the ideas held by Aristotle, was originated by **Johannes Philoponus,** in the first half of the sixth century. His teachings, known as the ‘ Impetus ’ Theory, maintained that bodies acquire their motion from an external force, and that they continue their motion in virtue of the impetus they have received. Fundamentally the term impetus implied, what to-day we should call kinetic energy, although the laws of energy were not known, at least as we know them to-day. In the ensuing centuries nothing was ever heard about this theory ; but when western civilization became acquainted with the teachings of Aristotle through the Arabs, this theory was revived almost at the same time. Even in 1277 had **Tempier,** a Bishop of Paris, rejected a whole series of Aristotle's laws concerning natural philosophy. In the Middle Ages, at the University of Paris, which was most famous at that period, views directly opposed to those of Aristotle were advocated, and formed the substance of lectures given by the leading professors.

William of Occam (at the beginning of the fourteenth century) and

Johannes Buridanus, one of his pupils (middle of the fourteenth century), together with their whole school, taught the Impetus Theory. Special reference must be made to **Buridanus**, whose views differed very considerably from those of the Ancient Greeks. He had even convinced himself that the heavenly bodies moved according to the same laws which were applicable to terrestrial bodies. Especially with regard to falling bodies had progress been made, for **Leonardo da Vinci** (1452–1519) and the Spanish friar, **Dominicus Soto** (1494–1560), had explained it as a case of a uniformly accelerated motion, even though an experimental proof was still lacking ; the second law of falling bodies, which expressed the relationship between distance and time, was known to **Nikolaus von Oresme** (*d.* 1382). At that time the University of Paris was the centre of the civilized world, and the works of its scholars were known everywhere. Galileo quotes **Occam** and other names, referring to them when he required support for his hypotheses. He had read their literature and was acquainted with their doctrines. But to Galileo must be accorded the credit for having established these laws beyond all doubt by making confirmatory experiments, for it was he who first made serious observations on natural phenomena and studied them in a spirit which, at that time, was quite original. Until the time of Galileo science had been the servant of philosophy. Galileo gave it independence, and this act science has rewarded by being subsequently of much greater service to philosophy than ever before.

But the problem of gravity remained a puzzle for a long time. The fact that it acts upon all bodies at the earth's surface was not understood for many years.

It was **Newton**, in 1687, who first enlightened us concerning these phenomena, when he showed that gravity is only a special case of an attractive force which exists between all bodies wherever situated. But alas, our knowledge concerning this most general of all natural forces is still restricted to our acquaintance with its mechanical effects, which occur upon the earth or amongst the heavenly bodies. That is the reason why it is always treated in physics as a branch of mechanics, though, strictly speaking, a section should be devoted to it just as much as to electricity, for example.

If several bodies are allowed to fall at the same time, light bodies such as down, feathers or flocks of wool being excluded, it is found that only very slight differences in the velocity of fall are exhibited. **Johannes Philoponus** advocated the following views, which were really much in advance of his times. He maintained that the effect of air resistance upon bodies of different weight was to retard them to considerably different extents, and that, in empty space, all bodies would fall at the same rate. If the experiment is repeated in a rarefied atmosphere, it is found that the velocity with which the lightest bodies fall approaches that of the heaviest bodies more closely, the more completely the retarding influence has been eliminated by the removal of the air. To-day, therefore, we are able to say, with absolute certainty and without exception, that in an air-free space all bodies fall with the same velocity. This is a very remarkable fact, especially when we consider what has been ascertained above concerning acceleration. According to that, forces proportional to the masses upon which they act must be present if the acceleration is constant. Since the force of gravity imparts to all bodies equal accelerations, it follows that gravity must act upon all bodies with a force which is proportional to the masses of these bodies. This is a very surprising result. Whatever the state of motion, we have for all bodies strict equivalence between the mass or resistance to acceleration on the one hand, and the force which attracts any terrestrial body towards the earth on the other. *Prima facie* it is improbable that this intimate connexion between mass and weight should be accidental. This has been further investigated, in order to discover whether, after perfecting the apparatus as much as possible, the equality between inertial and gravitational mass still remained. **Newton** placed different materials in the hollow bob of a pendulum and determined the periodic time. He was able to ascertain that the difference did not amount to one part in a thousand. In 1833 **Bessel** improved the apparatus still more, but the equality still remained, for he showed that the difference was certainly less than one part in six thousand. Finally, in 1891, **Eötvös** proved that the difference was even less than one in twenty million, probably less than one in a hundred million.

If such an equality is not accidental, we must seek in it a quite generally valid and fundamental law of Nature. Curious to relate, up to that time, nobody had made an experiment to ascertain this obvious connexion. It would be more correct to say that no result had yet been obtained concerning this relationship. For the repeated efforts to establish the real facts with certainty indicate, quite clearly, that the significance of the equality between inertial and gravitational mass has been noted from the beginning. In our own times, **Einstein** was the first person to formulate a hypothesis concerning motion, in which the equality between inertial and gravitational mass plays a decisive part. At the conclusion of this section we shall return to this point again.

From the fact that bodies fall with the same acceleration, it follows that, for the weights W and W' of two bodies of masses m and m' respectively,

$$W : W' = m : m' \qquad \ldots \qquad (30)$$

When a body is placed on a balance, we determine the mass which keeps it in equilibrium, and which must be attracted by the earth just as much as the body itself. But if $W = W'$, then must $m = m'$. In the first place therefore, the balance is used to ascertain when the weights of two bodies are equal. Since the two weights are equal, the equality of the masses is thereby deduced. In order to determine the two masses it is no longer essential for us to use the suspension railway, or any other similar contrivance ; the masses of a body can be determined much more simply and accurately with the aid of a balance.

The acceleration which all terrestrial bodies acquire when falling amounts roughly to 980 cm./sec.^{2}—this is denoted by g. The force with which a mass m is attracted by the earth is mg dynes, and the work which is performed against gravity when this mass is raised h cm. is $W = mgh$ ergs. Nevertheless, the acceleration due to gravity is not constant all over the earth. It changes with the altitude. The further one proceeds from the centre of the earth, the smaller does the acceleration of gravity become. Therefore, on account of the oblateness of the earth, g is smaller at the equator than at the poles. In addition to this, the centrifugal force at the equator opposes the earth's attraction at the equator more than in higher latitudes. The acceleration due to gravity, at the earth's surface, varies between 978 at the equator and 983 cm./sec.^{2} at the poles.

2. GRAVITATIONAL FIELDS OF FORCE IN GENERAL

After the stars had ceased to be regarded as bodies of a higher type, the fact that their movements are subject to the same general laws as apply to all other bodies could not have remained undiscovered for very long. The idea that bodies endeavoured to move towards the centre of the earth, which was then considered to be the centre of the universe, was replaced by that of **Johannes Philoponus,** who maintained that all bodies tended to unite themselves into a whole. This idea was extended by **Copernicus** (1473–1543) and **Kepler** (1571–1630) to include the tendency of rotating bodies to unite with the central body; *e.g.* the planets with the sun. But a more definite influence on these newer teachings concerning motion must be attributed to the Italian, **Borelli,** who, in 1666, first explained the motion of the heavenly bodies as due to a coordination between the attraction of the central body and the inertia of the revolving body. But he did not arrive at a solution to this problem by conjecture. An important step forward was made when **Huygens** discovered the law of centrifugal force [*cf.* p. 43]. This law, in combination with Kepler's third law, according to which the square of the periodic time for a planetary body is proportional to the third power of the major axis of its orbit, enabled the conclusion to be made, that the force of attraction is inversely proportional to the square of the distance. But the combined endeavours of the English philosophers, **Halley, Hooke,** and **Wren,** did not succeed in explaining the motions of the heavenly bodies. When they wrote to **Newton** about it in 1683, he replied that the problem had long since attracted his attention and that he had found a solution. Then in 1687 appeared Newton's famous work: *Philosophiae naturalis principia mathematica.* The falling apple, therefore, was not at all necessary to direct Newton's thoughts to the problem presented by gravity; the whole scientific world of that time was aflame with the desire to find the solution. But the ideas involved in the story of the falling apple were well suited to make people fully conscious of what lay at the bottom of the problem.

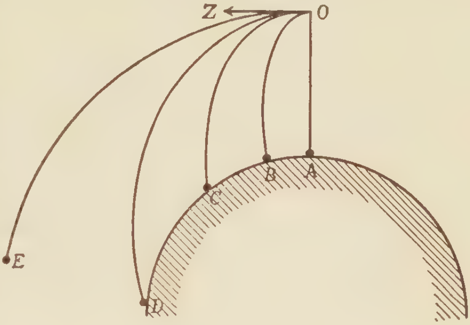


Fig. 19.

How far does gravity operate above the earth? If it brings down the apple from the topmost branch of the tree, and causes the stone to fall from the summit of a mountain, does it extend to the moon? If so, the moon would fall to the earth.—Then why does it not fall?—Or does it fall?—Is the attraction just equivalent to the circular motion of the moon in its orbit around the earth? If a stone is allowed to fall on its own accord, it will move towards the earth along the straight line OA, Fig. 19. But if, at the same time, it has a velocity in the direction OZ, it will move along the curved line OB; if the horizontal component is larger it moves along OC; and if the horizontal component of the velocity is sufficiently big it will move in a circle around the earth. It will always be falling towards the earth and yet it will never reach it. Exactly the same thing occurs in connexion with the satellites of the other planets, Jupiter and Saturn, and the motion of the planets around the sun.

From the beginning **Newton** saw that this was his main problem and he undertook to prove it completely. Since all bodies fall with the same acceleration, it is not necessary to know anything concerning the mass of the moon. Its path must be such that, every second, it is deflected by exactly that amount through which a body would fall towards the earth in the same time if it were in that position.

He assumed that the force would be inversely proportional to the square of the distance. This is, in fact, a most reasonable assumption for an effect which proceeds outwards from a point in all directions. Let, for example, *O* be a source of light, Fig. 20. The amount of light energy which is incident upon the surface *A*, will at a greater distance be distributed over the surface *B*. But since these surfaces increase as the square of their distance from *O*, the amount of light falling upon equal elements of area at *A* and *B* will be inversely proportional to the square of their distance from the source of light. The moon is at a distance from the earth which is approximately equal to sixty times the earth's radius, and the acceleration due to gravity at the earth's surface is 981 cm./sec.^{2} Hence the moon's acceleration towards the earth is $\frac{9810}{3600}$ mm. per sec.^{2} = 2·7 mm. per sec.^{2}; the distance through which it would fall in the first second of its motion is therefore 1·35 mm.

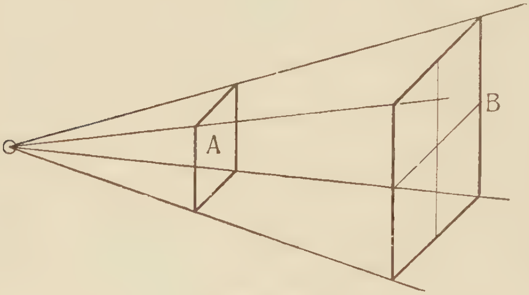


Fig. 20.

The radius of the moon's orbit amounts to 384,000 km., whilst its period of revolution is 27 days $7\frac{3}{4}$ hours, or $2{\cdot}36 \times 10^6$ secs. The distance traversed per second is therefore 1020 m. After one second the departure of the moon from the tangent [*cf.* equation (19), p. 42] is

$$e = \frac{v^2}{2r} = \frac{1{,}040{,}000}{2 \times 384{,}000{,}000} = 0{\cdot}00135 \text{ m.} = 1{\cdot}35 \text{ mm.}$$

Hence, at the end of the first second, the force which is directed towards the earth's centre is no longer directed along *AO* but departs from this by 0·5 second of arc *, so that the deviation from the assumed value must certainly be less than one part in a million.

The above calculation showed that the moon does actually move in such a way that it does always approach the earth, just as a stone would do, which is free to fall from that position. In addition, **Newton** proved that the motion of the planets around the sun, as expressed by the three laws of Kepler which were deduced from careful experimental observations, was an immediate consequence of his assumption that the planetary motions were regulated by a mutual attraction between these and the sun.

From the fact that all bodies are attracted to the earth, and that the earth, which attracts the moon, is likewise attracted by the sun, one was soon led to assume that all bodies experience a mutual attraction. This attraction governs the motion of a falling body, just as it does the motion of the heavenly bodies. The force is proportional to the masses of each of the two participating bodies. For, from the observation that all bodies fall with the same acceleration, it follows that the weight of a body is equal to its mass. But since both masses participate in the same way, the mass of the second body must occur as a factor in the law of attraction. If we denote by γ the force of attraction between two masses, each of 1 grm., separated by a distance of 1 cm., then the force between two masses m and m', separated by a distance r, is

---

* The circle is divided into 360° (degrees), each degree into 60′ (minutes), and each minute into 60″ (seconds).

$$f = \gamma \frac{mm'}{r^2} \quad . \quad . \quad . \quad . \quad . \quad (31)$$

An Explanation of the Law of Attraction. Newton imagined that even the smallest particles of matter attract one another, and that the enormous forces with which the planets are maintained in their orbits are due to attractions between the individual atoms which constitute the various heavenly bodies. Even though this force between two atoms at a distance apart of 1 mm. is so vanishingly small, and is very considerably reduced when the distance attains these enormous magnitudes (*e.g.* in the case of Saturn, which is at a distance of 1424 × 10^6 km. from the sun, the attraction is reduced 2 × 10^{30} times), yet none of this nebulous force is lost in the wide expanse of the universe. For the calculations concerning planetary motion which have been made on this basis have established the truth of this statement with the greatest accuracy.

Let A_1, A_2, A_3 . . . (Fig. 21) be the atoms of one body, B_1, B_2, B_3 . . . be the atoms of another body, and let the forces between them be represented by lines (one thinks of pieces of elastic joining the particles together).

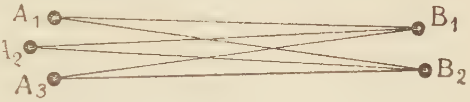


FIG. 21.

Each atom of A will exert upon all equidistant atoms in B an equal force, so that if the m atoms $A_1, A_2, A_3 \ldots A_m$ act upon the atoms $B_1, B_2, B_3 \ldots B_{m_1}$, mm_1 strings will be present. If we imagine that these cords become more feeble according to the square of the distance, we have a representation of the action of two bodies upon one another according to Newton's law.

A simple mathematical consideration shows that the true gravitational attraction due to a sphere of any magnitude is obtained if its whole mass is considered to be concentrated at the centre of the sphere. For this to be true, the mass must be distributed symmetrically throughout the whole sphere; it must therefore be homogeneous, or else concentric shells must have the same density at all points in them. In the original Newtonian law, as it is expressed by equation (31), this supposition has tacitly been assumed, since the law is expressed as though all the attracting particles were separated by the same distance r which separates the centres of the bodies. It is only in consequence of this supposition that the law expressed by (31) can be applied to masses of considerable extent, such as the sun, and planets.

Proof of the Formal Law of Gravitation. The Constant of Gravitation. The fact that any two bodies upon the earth's surface exercise a mutual attraction upon one another, given by the above law, was certainly always in **Newton's** mind. But he did not believe that such a small force could be successfully demonstrated in a physical laboratory. And yet such an experiment would have been of the greatest significance. For if, under some special circumstances, the force exerted between two known masses m and m' placed at a measured distance r apart, could be measured, the constant γ could be obtained from this result, and with it, the magnitude of the attraction between two bodies would be completely known.

In addition, many important conclusions could have been made concerning the masses of the planets. First, let us consider the earth. Since the acceleration of a falling body is due to the effect of gravitation between the earth and the falling body the centres of which are separated by a distance equal to the radius of the earth, we can write

$$mg = \frac{\gamma m m'}{r^2} \text{ or } g = \frac{\gamma m'}{r^2}.$$

Since g and r are known parameters in this equation, we can, as soon as γ is known, calculate from this equation the mass of the earth. If we assume that our earth is approximately a sphere of radius r, its mass is $4\pi r^3 d/3$, where d denotes the mean density of the earth. That is to say, it is such a density, that if it were the actual density at all points in the earth, the total mass as deduced from it would be the same as that which now results from the matter as actually distributed. The above equation could therefore be written

$$g = \gamma \frac{4\pi r^3 d}{3r^2} = \frac{4}{3}\pi\gamma r d.$$

The experiments which lead to a determination of the constant γ are really a determination of the earth's mean density. Numerous experiments have been made with an ever-increasing degree of accuracy, and these give 5·53 grm. per cm.^{3} as the earth's mean density.

Now since the bodies which are present at the earth's surface and alone accessible to us have, almost without exception, a density much less than this, it follows that much material having a density greater than 5·5 must exist in the earth's inaccessible regions. The general opinion is that the centre of the earth consists mainly of an iron core, the density of which is 7·8.

When γ is known, the equation

$$g = \gamma \frac{m'}{r^2}$$

allows the mass of any body having a revolving satellite to be found. (The sun and its planets, and the planets with their moons.)

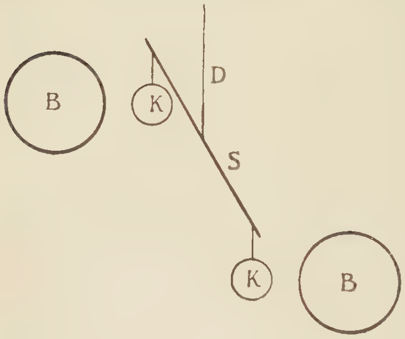


Fig. 22.

From the orbit of the revolving body, its acceleration g can be calculated, as we have already done in the case of our moon. Further, if astronomical observations have enabled r to be known, then the equation gives at once m', the mass of the rotating body, when the **gravitational constant** γ is known.

The first determination of the constant γ was made by **Cavendish** in 1798, more than a century after **Newton's** discovery. For that purpose he used only apparatus which Newton could easily have had at his disposal. His apparatus is now generally known in physics as the torsion balance. At the end of a long thin metal wire D, Fig. 22, is supported a long but light rod S carrying at either end a lead sphere K.

If such an apparatus is left undisturbed, it swings for a long time to and fro, finally coming to rest when the supporting wire is free from torsion. If the wire D is sufficiently fine, the sensitivity of such an apparatus becomes very great. The least breath of air is sufficient to deflect the balance from its zero position. It must therefore be enclosed in a glass case for protection against currents of air. When two large lead spheres of about 10 kgrm. mass are placed, one on the right of one of the small spheres, and the other on the left of the second one, the rod is thereby caused to rotate a little, until the torsional couple which arises in the suspension wire is sufficient to maintain equilibrium with the deflecting gravitational forces. Since the magnitude of the torsional force, as seen from equation (27), can be calculated from the time of oscillation, the magnitude of the gravitational forces becomes known. Moreover, it is easy to determine the masses of the spheres, and their distances apart, so that all magnitudes in equation (31), except γ, are known. Cavendish obtained 5·48 as the density of the earth, which corresponds to a gravitational constant $\gamma = 672 \times 10^{-10}$.

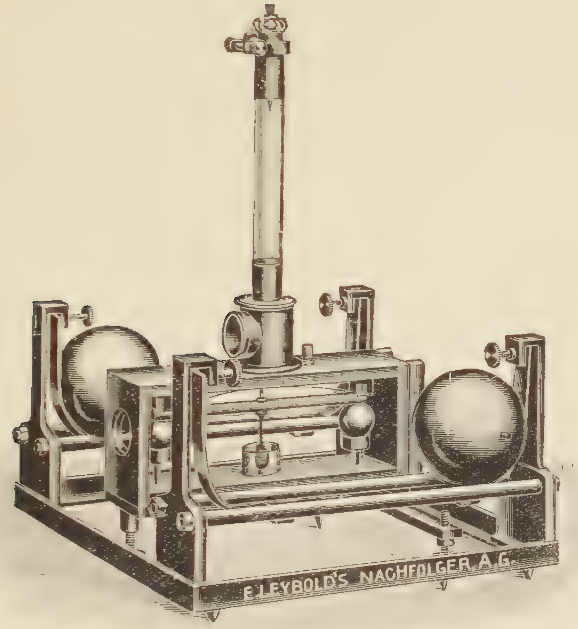


FIG. 23.

Later on the experimental methods were refined. **Boys** used, in 1894, in place of the suspension wire, a quartz thread, which he had discovered how to make. **Fr. Braun** placed the whole torsional balance in a vacuum, in order to reduce air disturbances to a very minimum. The values obtained by both these observers were concordant. The weighted average of all the results which have been obtained so far, gives for the gravitational constant the value $\gamma = 666 \times 10^{-10}$, which corresponds to a mean density $d = 5{\cdot}53$, in which it is only the last figure which may be uncertain to the extent of 1 unit.

The author has recently designed an apparatus, Fig. 23, which is so free from outside disturbances that the experiment can be used as a lecture demonstration. The large deflecting spheres run upon guiding bars, the one always moving left to right whilst the other moves from right to left, at the instant when the pendulum is deflected to a maximum, and just about to reverse its motion. The amplitudes become larger and larger, so that they are easily observed by a circle of spectators. Moreover, the instrument can be used to demonstrate how the attraction depends upon the masses and their distance apart, according to Newton's law, so that the gravitational attraction between bodies must now be included amongst facts which can be observed.

3. POTENTIAL

The study of the motions of bodies under the influence of gravity is one of the main branches of astronomy. A great number of these laws are summarized under the title, Potential Theory. This same theory, however, applies to many other forces, and therefore forms a very important part of the laws of moving bodies in general. We shall only discuss briefly those fundamental ideas which have to be used frequently in the succeeding chapters. Since gravitational forces exist all around the earth, and not only at its surface, we express this fact by saying that the earth is surrounded by a gravitational field of force. All material bodies which are present in this field of force experience an attraction, which is directed towards the centre of the earth. The force which is exerted upon unit mass placed at any point in the field is called the **intensity of the field** at that point. At a distance r from a planet of mass m this intensity is therefore equal to $F = \gamma \frac{m}{r^2}$, which is exactly as large as the acceleration with which bodies would move when placed in that position. The consequence of this is that work must be done when a body is moved away from the earth. Conversely, when the bodies move towards the earth under its attractive force, the earth does an amount of work which is equal to that done by the body in the first case. Since, however, the force varies with the distance, this work cannot be expressed simply, as the product of the force and the distance, since the appropriate value to be chosen for the force is not known. It is only when the body has moved over a short distance, in which the force due to gravity has not changed appreciably, that the work can be calculated as the product of this invariable force and the distance traversed. This is always done when g is put to represent the acceleration throughout the fall, and this g is assumed to be invariable throughout the motion. But more generally the force, at a distance r, is $f = \gamma mm'/r^2$. If the body moves to r_1, the force is there $f = \gamma mm'/r_1{}^2$. For an exact determination of the work done it is necessary to divide the path into small distances, so small in fact, that over each distance the force due to gravity can be considered constant. This is what we do in the Integral Calculus. If dr is such a small element of distance measured from the point r_0 outwards, the work done when the mass m^1 moves from r_0 to $r_0 + dr$ is $dW = \gamma \frac{mm'}{r_0{}^2}\, dr$. And if many such expressions are added together we obtain the work done in moving from r_0 to r_n. The Integral Calculus shows that the amount of work is

$$W = \gamma mm'\left(\frac{1}{r_0} - \frac{1}{r_n}\right).$$

This result can be obtained in the following manner without using the calculus. The mass m is to be moved outward from r_0 to r_1. The path which must be traversed is therefore $r_1 - r_0$, but the force is not known. If we assume that the force at r_0 is the force which is present everywhere, too great a value will have been used. If we assume that the force is everywhere the same as at r_1, then the value is too small. Thus $\gamma mm'/r_0^2$ is too large, and $\gamma mm'/r_1^2$ is too small. We therefore take a mean value, $\dfrac{\gamma mm'}{r_0 r_1}$, which will approximate more closely to the true value the smaller the distance $r_1 - r_0$ is chosen. Then the work is

$$\Delta W = \gamma\frac{mm'}{r_0 r_1}(r_1 - r_0) = \gamma mm'\left(\frac{1}{r_0} - \frac{1}{r_1}\right).$$

If we now proceed another small distance from r_1 to r_2, and then to r_3, r_4, . . . , the work done is found similarly in each case.

$$\begin{aligned}
\Delta W_1 &= \gamma mm'\left(\frac{1}{r_0} - \frac{1}{r_1}\right)\\
\Delta W_2 &= \gamma mm'\left(\frac{1}{r_1} - \frac{1}{r_2}\right)\\
\Delta W_3 &= \gamma mm'\left(\frac{1}{r_2} - \frac{1}{r_3}\right)\\
&\cdots\cdots\cdots\\
&\cdots\cdots\cdots\\
\Delta W_n &= \gamma mm'\left(\frac{1}{r_{n-1}} - \frac{1}{r_n}\right).
\end{aligned}$$

Adding these elements of work together, we obtain

$$W = \gamma mm'\left(\frac{1}{r_0} - \frac{1}{r_n}\right).$$

This is the same value as we obtained before by using the Integral Calculus.

Every movement of a material body through the earth's gravitational field of force is therefore associated with a performance of work, if, by such a movement, the distance from the earth's centre is varied. Only at an infinitely great distance does the force become zero. Practically the force becomes too small for it to be detected, by any means at our disposal, long before such a distance is reached. The maximum amount of work which it is possible to do, is performed when the body m' is removed to an infinite distance. The above formula shows this immediately, for we have only to put $r_n = \infty$, when its reciprocal becomes zero, and the work done when a body of mass m' at a distance r is removed quite out of the field is

$$W = \gamma\frac{mm'}{r}.$$

Now let us assume that $m' = 1$, in order to make the expression more simple. We find that the work which has to be done in order to remove unit mass from a distance r to one which is infinite, is

$$W = \gamma\frac{m}{r}.$$

This value is called the **Potential** V at the point r, and is written

$$V = \gamma\frac{m}{r} \quad . \quad . \quad . \quad . \quad . \quad (32)$$

For all points lying at a distance r from the attracting body this same amount of work must be performed; they are at the same potential. All the points having the same potential form a spherical surface of radius r, the centre of which is at the centre of the gravitational field of force. This spherical surface is called an **equipotential-surface.**

The force, which is always directed along a radius of this sphere, is everywhere perpendicular to this surface. From this it follows that no work is necessary to move a body over such a surface, just as a ship, moving on the surface of the sea, does no work against gravity.

But in order to proceed from a surface r_1 to a higher surface r_2, work must always be performed, and this is

$$V_1 - V_2 = \gamma m\left(\frac{1}{r_1} - \frac{1}{r_2}\right) \quad . \quad . \quad . \quad . \quad (33)$$

To whatever point on the other surface the body is brought, the amount of work required is always the same. One can pass in a radial direction from one surface to another, and then move over this second surface to the desired point without doing any further work, so that to reach any point r_2 on the second surface from any point r_1 on the first, the same amount of work is required. This amount is not only necessary, but it is essential, for there is no other path along which an amount of work different from the above can be done. For if the path is increased by moving in a direction inclined to that of the force, the component of force is diminished. The motion of a body upon an inclined plane is a simple example of the above. The plane can be inclined at any angle to the horizon, but the work done in raising a body is always equal to the product of the weight of the body and the height raised. [Equation (16), p. 38.]

Potential is not a property of any one body, but merely the result of calculation. Two facts form the basis from which the idea of potential is derived; first that of general gravitational attraction, and then the fact that work must always be expended whenever a material body is moved against the attracting forces. The idea of potential has been formed because it has proved a useful and convenient artifice in the solution of many physical problems.

The idea of potential can now be extended to bodies which are not quite spherical and to fields of force not having their origin at a point. For material bodies of other forms, the work which has to be performed when unit mass is taken from one point of the field to another right outside it is called the potential at the point. If the force is a continuous function of the position all points having the same potential must lie on a closed surface V_1, which is surrounded by other closed surfaces having a smaller or greater potential V_2. The force is everywhere perpendicular to these equipotential surfaces, and for a motion over such a surface no work is necessary; but in order to get from the surface V_1 to the surface V_2, an amount of work $\pm (V_1 - V_2)$ is always necessary.

If, in any special problem, it was known that the force F was constant along a path l, the work done could be expressed as

$$V_1 - V_2 = H \cdot l$$

from which it follows that

$$\frac{V_1 - V_2}{l} = H.$$

Since H represents the force on unit mass, $(V_1 - V_2)/l$ is the intensity of the field at the place in question. If the potential is not changing uniformly, it represents the mean intensity of field, but when the potential varies uniformly we have the true constant field strength.

When, under the action of gravity, a mass m has been brought from an infinite distance to a point at a distance r, an amount of work $W = \frac{\gamma m}{r}$ has been done by gravity upon this mass. Let us assume that this mass is a planet, which henceforth revolves round the central attracting body or sun at a distance r. It will then have a certain kinetic energy, due to its velocity. According to (21), the velocity v is related to the centripetal acceleration a by the equation $a = v^2/r$. But according to the law of gravitation $a = \gamma m/r^2$, so that $v^2 = \gamma m/r$, and the energy of a rotating body of mass unity is $\frac{v^2}{2} = \frac{\gamma m}{2r}$. This is only half as great as the energy expended by the central body in drawing the rotating mass into its position. During the approach the other half of the energy can be expended in a different way. If this did not happen, the rotating body would have received too great a velocity, and would not have been able to begin its motion around the central body in a circular path. According to its direction and velocity, it would have moved through the field of force, in a path represented by another conic section instead.

4. THE NATURE OF GRAVITATIONAL ATTRACTION

Newton had already raised the obvious question concerning the nature of gravitation. This he answered with his famous ‘ hypotheses non fingo ’, by which he wished to say, as the context shows, that the observed facts had, until then, given him no basis upon which a hypothesis concerning the nature of gravitation could be framed ; without one such experimental support he refused to make any conjecture concerning the nature of gravity. Unfortunately, we find ourselves in much the same position to-day. No efforts have been spared to unfold the puzzle presented by gravitation, but no other natural force has been so difficult to investigate quantitatively as gravity.

The difficulty is, above all, due to the smallness of the effects which can be produced in a laboratory. And in order to ascertain any small possible changes, the normal value must have been fixed with great certainty.

A further great difficulty lies in the general nature of gravitation. Whatever, therefore, one wishes to use as an experimental aid, also exerts an attractive force and thereby renders more difficult any definite assertion concerning any one point. Finally, there is, according to all experiments so far made, the extraordinary independence of gravity towards all possible influences which are likely to produce any alteration.

First and foremost we must remember that the general attraction between bodies is very different from the idea of force which we have formed on the basis of our own bodily experiences. Our hand only acts upon a body so long as it is in immediate contact with it. Gravity acts at a distance ; in fact distance does not limit its activity ; it not only brings the apple from the tree, or causes the avalanche, but it acts upon the moon and confines it to its path. Is it possible at these great distances for these forces to appear out of nothing ? Such a view can never be conceived. The ‘ actio in distans ’, *i.e.* the idea of action at a distance, without the intervention of something tangible, is a conception which cannot satisfy our intellectual powers and is therefore refused by all. **Newton** himself held this view. It is true that his own law states nothing further than the magnitude of the action at a definite distance. But injustice is done to Newton if, from the above, it is concluded that he wished to imply that no intervening medium was necessary. His law only states something concerning the magnitude of the force, and says nothing about its manner of propagation. At the end of his treatise on mechanics he stated explicitly, that he made no additional assertions concerning the manner in which gravity acted at a distance, because he had not yet succeeded in deducing from the phenomena the reason for this property of gravity. He also speaks of an extremely light something, which he calls 'spiritus quidam subtilissimus', and which, to-day, is called the **Æther**. It penetrates all objects and is the medium by which all bodies are able to exert a mutual attraction upon one another even when they are only at a very short distance from one another. But sufficient experiments were lacking for more to be said. By these statements, Newton meant to imply the action of the intervening medium, and it is this action which we must proceed to investigate. For if there is no direct action at a distance, the space between two attracting bodies must behave differently when the bodies are present, than when they are not there. All experiments in this direction have failed.

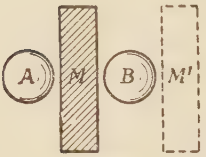


Fig. 24.

One would have thought that the influence of an intervening medium would have shown itself in the following way; the effect of two bodies A and B, Fig. 24, upon one another would be different according to whether the intervening space was filled with one material or another, and especially if at one time it were exhausted and afterwards filled with a heavy material. Such experiments would be very difficult to perform because the ponderable media would exert an attractive force which would be simultaneously present.

Moreover, Newton's law itself asserts that the material which fills the intervening space does not affect the force of attraction. For there are only a small number of mass particles actually lying on the earth's surface which are able to exert an attraction upon the moon directly through free space; the remainder must more or less act through the layers of earth which extend in front of them. If the medium did influence the force of attraction, then the law of gravitation would have been considerably altered, probably by a factor indicating the variation with position. Thus all the questions which have been put to Nature on this matter have hitherto been answered in the negative, or have not been answered at all.

The influence of an intervening medium would have at least been suggested, if it could have been shown that the propagation of gravitation required a certain time. But here it has not been possible to devise direct experiments, because from the beginning it has not been clear what meaning is to be given to questions concerning the velocity of propagation. If the question is raised in light or acoustics, a definite meaning is attached to the time in which an emitted signal is reflected from a distance mirror and returns to the sending station. The distance and time can be accurately measured. But if the attraction between the earth and the sun is mutual, does the effect originate at the sun or at the earth? It therefore only appears possible to speak of a velocity of propagation, when the coordinates of place and time for the signal can be specified. In the present problem neither is known. Finally the question has been considered by assuming that two bodies are initially at a very large distance apart. If it is now conceived that they suddenly approach one another, is the attraction from the first instant that corresponding to their closest distance, or does time elapse before the attraction attains its new value? This question can be answered by considering the motions of the heavenly bodies which move in ellipses having a large eccentricity. In such a case the force of gravity will change considerably on account of the varying distance of the body from its sun. The path of such a body can be calculated first on the assumption that the attraction at every point immediately attains its final value, and then it might be assumed that there is a certain interval of time before the new value is reached. Since small deviations have always to be considered in such a problem, it cannot be claimed that these calculations yield a final solution, especially since different astronomers have obtained different results. In 1805 **Laplace** obtained a value for the velocity which is 10^7 times that of light; **Hepperger** in 1888 was satisfied with a value which was 500 times the velocity of light; in 1912 **Oppenheim** required a value for the velocity with which gravity was propagated which was at least twelve million times that of light, whereas **Gerber** maintained that the velocity was equal to that of light; taking his calculations for a basis, he believed that the medium whereby the force of gravity was propagated was identical with that for light. However, it is seen from the divergency of their assertions, that the astronomers themselves do not regard any one of these values as final, since there is no unanimity even as to the order of magnitude of the velocity of propagation of gravitation.

A few further remarks must be made in order to obtain an indication of the effect due to an intervening medium. The fact that the force is inversely proportional to the square of the distance follows immediately from the three-dimensional nature of the space in which we live. But if the medium influences in any way the propagation of this force, then it can be conceived that it will be more or less reduced in intensity as it travels along, just as the intensity of light is weakened when it traverses the earth's atmosphere. Such an effect would cause the force to diminish more rapidly than one which varies as the square of the distance. It would therefore be necessary to consider whether a law with $\frac{1}{r^{2+\delta}}$ instead of $\frac{1}{r^2}$, or some similar law, represented the planetary motions better than Newton's law. All variations which have been proposed by different investigators, however, are so extraordinarily small, that it is quite legitimate to doubt whether, on the whole, they represent any improvement on Newton's law.

If we wish to summarize, it must be made quite plain, that all attempts to demonstrate any effect whatsoever, which is attributable to the medium, have resulted in complete failure.

Finally therefore, without having any experimental support, an attempt has been made to form some conception of a possible way in which gravity could be propagated through space. **Isenkrahe** has collected the above data in a small book (*Das Rätsel der Schwerkraft*, Braunschweig 1879). Naturally not one of these proposed theories makes any claim to be considered as being correct.

5. INERTIA AND GRAVITATION

The problem of gravitation has, during the last decade or so, been considered from quite a new aspect by **Einstein**, who, in his General Theory of Relativity, has found a relationship between gravitational and inertial mass. These attempts are very attractive because they solve two problems at once. Until then inertia had been considered as something final. Since it appears to be inseparable from every atom and is not influenced by any external agency whatsoever, it is quite obvious that it must be considered as a characteristic constant of matter. More than that we cannot say.

We have already remarked on p. 49 that the inertial mass of a body is intimately connected with its gravitational mass, and that the equality of the two corresponding forces is so perfect, that any difference which might exist is beyond the sensitivity of our present-day instruments. Einstein regarded this conformity as an indication of an innate connexion between the two phenomena.

In these ideas he was really anticipated by **Ernst Mach** (1838–1912). In 1872 Mach emphasized the facts that we could only recognize the motion of a body relative to another, and that all actions between one body and another likewise depend upon a **mutual** action between the two bodies. He therefore considered it to be very remarkable if the effects due to inertia should form an exception to these laws. When bodies rotate, a centrifugal force is called into play : our earth is made oblate, whilst wheels, which revolve too rapidly, fly in pieces. This effect appears to indicate that a rotatory motion is absolute. When **Newton** caused a glass vessel containing water to rotate by means of a cord from which it was suspended, the centrifugal force did not manifest itself so long as the glass alone rotated, and the water did not participate in this motion. The motion was only relative. It was only when the water rotated that it rose up the sides of the vessel in consequence of the centrifugal force. **Mach** pointed out very emphatically that it is an overhasty conclusion, if, from this and other similar phenomena, we deduce that centrifugal force only depends upon **absolute** rotation and permits us to infer certain things about absolute motion, since we do not know what part surrounding bodies play in connexion with the rotation. We cannot even say whether or not centrifugal force would arise in the absence of these surroundings. Nobody has yet made such an experiment. Mach imagined that phenomena due to inertia are just as much a reciprocal action between a moving body and its surroundings, as those due to gravitation.

But inertia only asserts itself when one endeavours **to change the relative motion of a body**, whereas gravitation always exists as an attraction. In both cases, however, one factor would be respectively the mass of the attracted body and of the inertial body, and hence the equality of weight and inertia would be a simple deduction. But Mach was able to say nothing about the other factor on which inertia may depend. It is certainly not a single neighbouring body, as in the case of weight, which depends on the earth and not on an influence of Sirius ; or, as can be proved by the torsion balance, the action on the oscillating mass is very markedly determined by the neighbouring heavy lead spheres. Mach indicated two possible agents, either the rest of the material world, or else a medium in which the motion took place. We should now call this medium the æther. According to Mach, we may therefore regard the inertia of a body as a relationship between an accelerated body and the world around it.

This relationship makes itself felt whenever the velocity of a body is changed with respect to the rest of the world, *i.e.* according to the second postulate, with respect to the æther. Centrifugal forces, for example, would therefore arise if the whole world were rotating about one particular object, just as when that body rotates with respect to all other surrounding objects. We have therefore to conclude that centrifugal actions are merely relative. All this, of course, could be tested by experiment. If a vessel containing mercury is rotated the liquid rises against the walls in consequence of the centrifugal force. But if the vessel were now surrounded by a thick mantle of steel which was then set in rotation, according to **Mach** the mercury would still rise, whereas **Newton** would maintain that its surface would still be horizontal. Although such an experiment is quite conceivable, actually such a large mass cannot be caused to rotate, or else if a smaller body is rotated the apparatus, which we have at our disposal, is not sufficiently delicate to detect any effect which may be produced. If, therefore, we cannot make a decision between Newton and Mach, we are no longer able to regard the statement, that the inertia of a material body is something belonging only to that body itself, as one founded on definite proof. The opposite view is also possible, and the fact that there is equivalence between the gravitational and the inertial mass of a body has always been in favour of it.

Einstein has adopted these ideas of Mach and used them in his theory of relativity. Since then Mach's ideas have not only been a subject for much consideration, but they have been extended by workers in many different directions. However, if the attempt is made to carry through Mach's ideas to completion, many difficulties arise, so that for the present we have to regard the question as unsolved.

CHAPTER IV

ELASTICITY

As solid bodies (in contrast to liquids and gases) we designate those bodies which have a definite and independent form, so that we can recognize them again from this shape. The rigidity of all bodies is, however, only relative; absolutely rigid bodies would be those whose shape could not be altered even by an infinite force. Such a body is said to be perfectly rigid, but such bodies do not exist, for all actual bodies suffer a more or less slight deformation when they are acted upon by forces which tend to change their shape. The type of deformation is, however, very different for different materials. Above all it depends upon the magnitude of the force; only moderate forces are necessary to leave an observable impression on lead, but these same forces leave much smaller impressions on soft iron. Then again, the deformation depends upon the kind of material which is being examined. After a leaden sphere has been dropped upon a smooth tile the flattened area which is a result of the force necessary to bring it to rest is clearly visible. If an equally massive steel or glass sphere is allowed to fall from the same height, no such mark is to be found upon the sphere. If, however, the tile is blackened with graphite, or some similar material, it is at once seen that the glass sphere was quite noticeably flattened, for it receives a circular black mark to indicate the points at which contact occurred. These last-mentioned bodies, which, after having been flattened by impact, recover their former shape as soon as the external forces acting upon them are removed, are said to be elastic bodies; those bodies of the first type are said to be inelastic. Although it would appear as if these two types were mutually exclusive of one another, all actual bodies possess something of both, but in very different degrees. For very small forces all bodies are elastic, even perfectly elastic; *i.e.* they assume their former shape completely without retaining any trace of an impression which may have occurred. The limit up to which a depression is completely effaced is called the **elastic limit of the material.** Different bodies possess elastic limits which are very far removed from one another. It is just the different positions of the elastic limit which differentiate elastic and inelastic bodies from one another. In the case of lead, for example, a permanent trace is left even after quite minute impressions; for steel and glass, however, much greater impressions are completely effaced. But even for these substances there is a limit, above which, the return to the original shape does not occur. Hence all bodies can be considered as perfectly elastic so long as one is careful not to go beyond their particular elastic limit.

If a body is loaded beyond its elastic limit, it loses its shape at an ever-increasing rate as the load is increased, until, for an additional increased load, it finally breaks. If there is a marked alteration in shape before the rupture occurs, the body is said to be tenacious and ductile; if the rupture happens without any very visible change in shape after the elastic limit has been reached, the body is said to be brittle.

In industry there is a great demand for elastic springs; in architecture, too, especially in the construction of bridges, it is very important to consider the elastic properties of the materials being used.

Certain types of deformation are also very necessary for an understanding of physical processes, and must therefore be discussed briefly here.

1. **Longitudinal Elasticity.** Let us assume that we have a long body, Fig. 25, in the form of a steel wire, iron band or a long piece of india-rubber tube: the upper end is securely fastened whilst the other one carries a pan into which weights may be placed. A vertical scale enables the length of the body to be ascertained when necessary. Before being loaded let L cm. be the length of the object. If a weight is placed in the pan, the latter sinks because the wire extends under the influence of the weight. The extension is l cm. let us say; a doubling or trebling of the weight produces an increase which is double or treble the former amount, the increase in length always being proportional to the stretching force (so long as the elastic limit is not exceeded).

The extension takes place throughout the length of the whole wire and not at the expense of any one particular piece. The process is something as follows: any two cross sections of the wire which before the application of the load were at a certain distance apart, are now a little further apart. At first the lowest layer gives way a little under the application of the force and then exerts a pull on the layer immediately above. This in turn yields under the force to which it is now subjected, until it too is pulling on the next layer with an equally large force. In consequence of this the third layer yields and so the whole process is repeated until the point from which the wire is supported is reached so that no further yielding can take place. Now whenever the distance between two layers in the wire is increased the pan which carries the weights falls by a similar amount so that the total extension is equal to the sum of the elongations of the individual layers. Since, for wires of the same material the extension for every unit length takes place irrespective of the actual length of the wire, it follows that the total extension of the wire under the same load is proportional to the length of the wire, a result which can easily be verified experimentally. If we desire to compare the elasticities of different bodies, we must thus always compare bodies of the same length with one another. We always ascertain, therefore, the extension of one centimetre of the material, and call it the specific extension λ. We therefore have

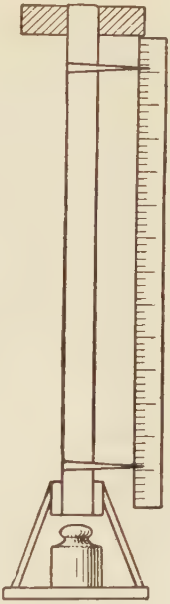


FIG. 25.

$$\frac{l}{L} = \lambda \qquad . \quad . \quad . \quad . \quad . \qquad (34)$$

As regards the relation of this to the stretching force, it is not immaterial whether the force is applied to a thick or thin wire, a fact which can easily be established as follows. We will assume that there is a series of wires all of the same material and length, each having a cross section equal to 1 mm.^{2} If one of these wires is extended by means of a 1 kgrm.-wt. there is a definite extension, which is identical for the second and all the remaining wires. If ten wires, let us say, each hanging under the influence of its own attached weight, are now bound together to form a wire rope, neither the tension in each strand, nor its length, will be altered. The same extension takes place when the wire of 10 mm.^{2} section is stretched with a weight of 10 kgrm. as when a 1 mm.^{2} wire is loaded with 1 kgrm. If, therefore, we wish to express an opinion concerning the behaviour of different materials, wires of the same cross-sectional area must always be used for the comparison. If a is the cross section and F the force applied over this whole area, then the force which is operative on unit area (1 mm.^{2}) is $f = \frac{F}{a}$. In the sequel it will always be supposed that a 'wire' of 1 cm.^{2} cross section has been stretched by the force f, since we know that for other areas the same extension will occur if, instead of f, we write $F = fa$ in the calculations.

Since the total increase in length, l, is proportional to the stress, it follows that the specific extension, λ, will similarly be proportional to it. Thus, the ratio

$$\frac{f}{\lambda} = e \qquad (35)$$

is constant for one and the same body. This constant e is therefore large when a large stress f produces a small elongation, *i.e.* when the material is only deformed with difficulty. This constant is called the **elastic coefficient** or **modulus of elasticity**. Considering the above relationships, $f = F/a$ and $\lambda = \frac{l}{L}$, we have

$$l = \frac{f}{\lambda} = \frac{FL}{al} \qquad (36)$$

or,

$$e = \frac{FL}{ae} \qquad (37)$$

Hooke's Law (1645–1703). The law $e = \frac{f}{\lambda}$ can be expressed as follows : the modulus of elasticity is that stress which will cause the specific extension λ to be equal to unity ; for when $\lambda = 1$, we have $e = f$. Since there is scarcely any material which allows such an elongation to take place, this definition is best avoided, especially since this can be done quite easily, as we have shown above.

The laws which were there expressed are not only applicable to the extension but also to the compression of bodies, which are so shaped that they are not thereby buckled.

In order to produce an elongation, the force, which is applied to one end of the wire, has to traverse a distance l so that the extension of a material is always intimately connected with a performance of work. The force, however, is not constant ; at first it is very small but grows, as we have seen, at a rate proportional to the elongation. In such a case it may be assumed that, instead of a continually increasing force, the average force has been effective throughout the total distance. If F is therefore the force required to produce an elongation l, the average force is $F/2$, so that the work done is

$$W = \frac{Fl}{2} \qquad . \quad . \quad . \quad . \quad . \quad (38)$$

2. **Elasticity due to Bending.** This particular modulus of elasticity is very closely related to the elasticity due to bodies in tension. If a rod or bar is clamped at one end, Fig. 26, and loaded at the other, the rod is bent downwards. Similar effects are observed when a bar is clamped at both ends and loaded in the middle. On account of this bending, one part of the body bulges outwards and extends, whilst the other is compressed, until the elastic forces which result from this are sufficient to maintain equilibrium with the external forces acting on the body.

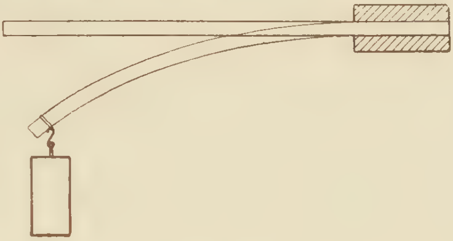


Fig. 26.

3. **Torsional Elasticity.** If a thin wire, clamped at its upper end, and carrying a horizontal bar at the other, is allowed to hang freely from the ceiling, or some sort of cross beam, the wire can be twisted about a vertical axis by rotating the bar in a horizontal plane; since the upper end of the wire is fixed, the wire must undergo a change of shape, which is caused by every layer of the wire being slightly rotated with respect to the one immediately above it; this, too, rotates with respect to the next upper layer, the process being repeated until the topmost layer is reached, and this, being fixed, cannot rotate. If we measure the angle of twist in circular measure the force which produces the twist is again proportional to the angle through which the lower end of the wire twists. If D is used to denote the moment of the forces required to produce a twist whose circular measure is unity, then the moment necessary to produce a twist φ is given by $F = D\varphi$. The constant D is called the 'restoring force' of the system. It is the magnitude of the force which arises when the wire is twisted through an angle whose circular measure is unity.

If the above paragraphs are read once again it will be noticed that they contain only facts of observation or else simple deductions from these. No attempt has been made to explain the phenomena nor to account for the forces which, for one body, cause it to recover its original shape completely, even after a large deformation has occurred, whilst for another, the original shape is not so regained. It is easily seen that elastic phenomena must be due to forces which tend to hold the small particles of a solid body in a definite position with respect to one another. The explanation of elasticity must ultimately depend upon the relationships between the atoms of that structure which we call a solid body. It is unfortunate that our ignorance of this branch of science is so great, that whilst we may ascertain the facts correctly, yet we can offer no explanation of them which will correlate the facts with the forces which exist between the atoms and molecules in the body. It is only in very recent times that experiments, having this as their ultimate aim, have been made.

4. **The Modulus of Bulk Elasticity.** In addition to the elastic coefficients which we have already mentioned, there is the modulus of bulk elasticity. It is associated with certain bodies that can be com-

pressed into a smaller volume and then exhibit characteristics which are similar to those due to elasticity. The volume elasticity is a property common to all bodies but it has an especial significance with gases. Only the essential facts concerning this property will be mentioned.

Suppose that we have a long cylinder (Fig. 27) closed at its lower end, whilst a moveable piston, which can be loaded at will, moves inside the cylinder. The cylinder is filled with air at atmospheric pressure. The force F, which is distributed over the whole area a of the piston, can be considered replaced by smaller forces f, each operating upon 1 cm.^{2} of the piston, where $f = F/a$.

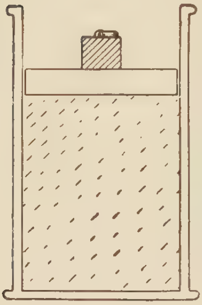


FIG. 27.

Let V be the initial volume of the cylinder. If weights are placed on the moveable cylinder so that the force operative on every 1 cm.^{2} is f, the piston sinks and the diminution in volume, v, is easily derived from h, the displacement of the cylinder, for $v = ah$. Since all layers of the gas are compressed uniformly, the total diminution in volume will therefore consist of the sum of the small changes taking place in each element of volume, and under the same conditions the diminution v will be greater the greater the volume V of the gas which is under compression. We may therefore conclude again that the diminution for unit volume is $v/V = \lambda$. Now experiment has shown that, over a wide range, this specific compression, λ, is proportional to the force producing the compression. As before, it is appropriate to call this ratio the coefficient of elasticity. The equation

$$E = \frac{f}{\lambda} = \frac{FV}{av}$$

is therefore true now. If we write $V = aL$ and $v = al$, which is permissible only for a cylinder, the a vanishes from the equation, and we have an equation which is equivalent to (36), viz.,

$$E = \frac{FL}{al}.$$

It should here be noticed, however, that a deeper insight into the inner mechanism of a gas has been obtained than in the case of solids and liquids. For we know that the phenomena arising from the volume elasticity of gases can be explained in terms of the impacts made by the gaseous molecules upon the walls of the containing vessel; in fact, the law deduced above is like **Boyle's** Law, nothing more than an expression of a relationship between the pressure and volume of a gas at constant temperature. In fact, Boyle's work was first published under the title, *The Elasticity of the Air*. This is a subject about which more will be learnt at a later stage.

CHAPTER V

IMPACT

THE events which take place during the collision of two bodies furnish a very good example of the problem which has already been mentioned, viz. when a force is only operative for a very short interval of time, and during this time any changes which occur in them cannot be traced; it is only the total change after collision that can be observed.

We assume that two bodies, of mass M_1 and M_2, Fig. 28, which are completely free to move in any direction, have velocities v_1 and v_2 ($v_2 > v_1$). Let both be moving in the plane of the diagram and suppose that M_2 suddenly hits M_1, so that it cannot continue its path without changing the path of M_1. What will happen?

In general, that will depend essentially upon the properties of the two bodies. Let us assume that both bodies are perfectly rigid. Then both, or at least one of them, must *instantly* change its state of motion. An infinitely great force is necessary for this to occur. Since action and reaction are equal, both bodies would be acted upon by infinite forces, the effect of which would at first only be felt by the particles immediately in contact. If the different constituent parts of the body are held together by infinite forces, then this effect would be imparted immediately throughout the whole material. Such a state of affairs would cause our world to be in a sorry plight, for such infinitely rigid materials could neither be worked, nor cut up for our use. If the particles are not bound together with infinitely great forces, however, they would fly to pieces even after a most gentle collision. (It should be noticed that a perfectly rigid body does not necessarily imply that its molecules are held together with infinitely great forces.) Fortunately, however, there are no perfectly rigid bodies. Hence all bodies after impact with others will undergo at first a slight flattening at the points which are mutually disputing the position. M_2 is able to advance by the small distance through which the extreme edges of the spheres yield (from A to B in Fig. 28). During the corresponding interval of time both spheres are able to change their state of motion so that it is not necessary for this to occur instantaneously (the flattening of a lead sphere, or the black disc upon an ivory ball which is allowed to drop upon a blackened surface, may be mentioned). During the time of impact both bodies act upon each other. If a perpendicular is erected through the centres of these small surfaces, this perpendicular indicates the line of action of the forces. If the line passes through the centres of gravity of the bodies, the collision is said to be central, otherwise it is called

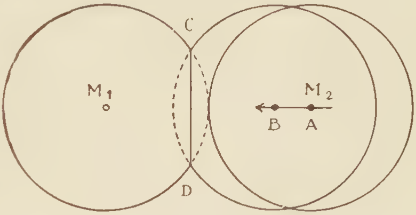


FIG. 28.

eccentric. In spheres these perpendiculars always pass through the centres, which points are also their respective centres of gravity if the spheres are uniform in density. Moreover, if this direction coincides with the line of motion of the colliding spheres the collision is said to be direct; otherwise it is referred to as being oblique.

It is owing to the forces which operate in the given direction that the faster body is retarded whilst the other is accelerated. The change in momentum, however, is the same for each body. If u_1 and v_1, and u_2 and v_2, are the respective velocities of the two spheres at the beginning and end of any time interval during which the mutual action still persists, then equation (12) [*cf.* p. 35] is always valid; *i.e.*

$$m_1\ (u_1 - v_1) = - m_2\ (u_2 - v_2).$$

This mutual action only persists until the velocities, in so far as they occur in the direction of impact, have become equal.

Inelastic Impact. The subsequent behaviour of the two spheres will depend upon whether the existence of the small flattened portion *CD* gives rise to any further effects or not. Let us assume first of all that this is not so. The two bodies are flattened and remain thus. Such spheres are completely inelastic, as, for example, spheres of lead, plasticine, glazier's putty, etc.

Since the above equation is applicable for every instant during the impact, it is necessarily true for the last instant of impact. But it is known that, at this instant, the two velocities are equal so that we have a second equation in addition to the above, viz. $u_1 = u_2$. If, for the sake of simplicity, this common velocity is denoted by u, this value can now be calculated for

$$m_1(u - v_1) = - m_2(u - v_2)$$

so that

$$u(m_1 + m_2) = m_1v_1 + m_2v_2$$

or

$$u = \frac{m_1v_1 + m_2v_2}{m_1 + m_2} \qquad . \quad . \quad . \quad (39)$$

As a limiting case we may mention that in which impact occurs between an inelastic body and one which is infinitely large; for our purpose the impact of an inelastic body upon a solid wall rigidly fixed in the earth may be regarded as such a collision.

If m_1 denotes the mass of the infinitely large body, we may divide both numerator and denominator of the above by m_1. Now the fraction $\frac{m_2}{m_1}$ is negligible compared with any finite value and so cannot affect u; it may therefore be called zero, and we have

$$u = \frac{v_1 + v_2\left(\frac{m_2}{m_1}\right)}{1 + \frac{m_2}{m_1}} = v_1.$$

an equation which tells us that, after impact, an inelastic body remains in contact with the solid wall.

Elastic Impact. The equation (12) is also true for elastic collisions, and since the action is always a mutual one, the equation is also true at every stage of the impact: therefore it is also true for the final velocities. But we can no longer say that these end velocities are equal to one another.

For, the velocities having become the same during the time that the bodies were being deformed, the colliding bodies henceforth begin to act upon one another on account of the deformation which has occurred. Now elastic bodies are characterized by their desire to regain their original shape after having been deformed: perfectly elastic bodies are those which do regain their former shape completely. But when the two bodies, which are at the present time moving forward with the same velocity, begin to regain their original size, a pressure is exerted between them: m_2 will be retarded, whilst m_1 is accelerated. For bodies which are perfectly elastic, the pressure at any instant during the time that they are thus exerting a mutual influence is exactly as large as the pressure was at each corresponding stage when the deformation was taking place. The same changes in impulse will therefore occur again, but in the reverse order. Consequently the change in velocity for each body, during the second half of the collision, is exactly equal to that which took place in the first half.

Let u be the velocity common to either body at the end of the first half of the impact, so that m_2 has its velocity diminished by $v_2 - u$, whilst m_1 increases its velocity by $u - v_1$.

In the second half of the collision these same changes again take place so that, on the whole, the change in velocity for m_2 is $-2(v_2 - u_2)$, whilst for m_1 it is $+2(u - v_1)$. Hence their final velocities are given by

$$u_2 = v_2 - 2(v_2 - u) = 2u - v_2$$

and

$$u_1 = v_1 + 2(u - v_1) = 2u - v_1.$$

Introducing the value for u from equation (39), we have, after rearranging the expression,

$$\left.\begin{aligned} u_1 &= \frac{2m_2v_2 + v_1(m_1 - m_2)}{m_1 + m_2} \\ u_2 &= \frac{2m_1v_1 + v_2(m_2 - m_1)}{m_1 + m_2} \end{aligned}\right\} \quad . \quad . \quad . \quad (40)$$

These formulae, which are quite general, enable us to examine several simple but important cases. If one sphere is initially at rest, *i.e.* $v_1 = 0$, we have

(*a*) if $m_1 = m_2$, $u_1 = v_2$, and $u_2 = 0$;

(*b*) if $m_1 \gtrless m_2$, $u_2 \lessgtr 0$ according as $m_1 \gtrless m_2$.

These equations show that when impact occurs on a smaller but stationary sphere, m_2 will travel with less velocity but in the same direction. If the spheres are equal, m_2 will be brought to rest, whilst when a smaller body collides with a larger one, the smaller body will acquire a retrograde velocity. If, in this latter example, the moving body encounters an elastic but infinitely large wall, it will return with its original velocity, *i.e.* $u_2 = -v_2$.

The problem concerning the collision of perfectly elastic bodies can also be solved by using the principle of the conservation of energy. Since the two bodies acquire their original shape again after the impact—no trace of the deformation being left—no energy can have been converted into some other form. It therefore follows that the total kinetic energy after impact must be equal to the total kinetic energy before impact, *i.e.*

$$\tfrac{1}{2}m_1v_1^2 + \tfrac{1}{2}m_2v_2^2 = \tfrac{1}{2}m_1u_1^2 + \tfrac{1}{2}m_2u_2^2.$$

Writing this equation in the form

$$m_1(v_1^2 - u_1^2) = -m_2(v_2^2 - u_2^2)$$

we obtain $$m_1(v_1 + u_1)(v_1 - u_1) = -m_2(v_2 + u_2)(v_2 - u_2)$$

Now according to equation (12)

$$m_1(v_1 - u_1) = -m_2(v_2 - u_2)$$

so that, by division, we have

$$v_1 + u_1 = v_2 + u_2, \text{ or } v_1 - v_2 = -(u_1 - u_2) \quad . \quad . \quad (41)$$

The values for u_1 and u_2 can now be obtained without much difficulty.

The former proof, however, is to be preferred since we have not yet proved that the energy law is universally true. Moreover it is not obvious whether or not some of the energy might have become converted into some other form, *e.g.* some of it might have become thermal energy. It would therefore be more satisfactory if one showed that the values given in (40) were consistent with the assumption that the kinetic energies before and after impact are equal: this can be done very easily by inserting the values for u_1 and u_2 in the expression

$$\tfrac{1}{2}m_1u_1^2 + \tfrac{1}{2}m_2u_2^2.$$

Equation (41) is worthy of note since it proves that the difference in velocity between two perfectly elastic spheres before impact is equal to the difference which exists after impact, but with its sign changed.

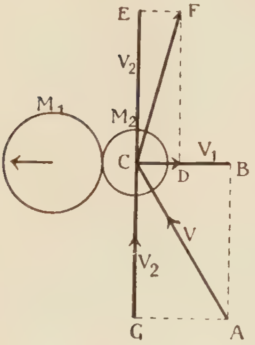


Fig. 29.

In the above paragraphs we have derived the laws of impact for bodies which are either perfectly elastic, or perfectly non-elastic. In Nature bodies of neither type are to be found; but there are many substances which approach these limiting cases so nearly, that these laws can be studied experimentally, and thereby confirmed.

Oblique Impact. In studying this type of collision the velocity $AC = V$, Fig. 29, is resolved into two components; the one, $BC = V_1$, lies in the line joining the centres of the two spheres, whilst the other $GC = V_2$ is perpendicular to it. The components in the former direction will be altered according to the laws of impact already established—let CD $(= u_2)$ be the final component velocity along CB. The other component is unaltered so that after impact the resultant velocity of M_2 is CF. If M_1 had been moving before impact it would have been necessary to resolve its velocity into two components, and then treat the problem on lines similar to the above. When impact occurs between a sphere and infinite wall, both of which are perfectly elastic, then the sphere is reflected from the wall with a velocity equal to that which it originally possessed but in such a direction that the angle of incidence (the angle between the normal and the incident path) is equal to the angle of reflexion.

CHAPTER VI

WAVES

1. Fundamental Ideas

IN Chapter II (p. 44) it has been shown that a particle executes a normal harmonic motion whenever it is constrained to its position of rest in such a way that when it is displaced even by the smallest amount from this position, a force arises which tends to pull it back again into its original position. This force increases in magnitude as the displacement becomes more pronounced. In Nature there are not only individual particles, but also whole bodies, in fact, bodies of quite considerable magnitude, in which these properties are manifest. All the individual points of such a body are capable of executing a normal harmonic motion.

These bodies which occur in Nature, however, are generally so contrived that the forces which tend to hold one particle in its position, are the same forces which exert their influence upon the neighbouring particles. The position of rest is not fixed in space but is only a position relative to the other particles of the same body. Let us consider the problem of a body which consists of a series of inert masses bound together by means of a rubber band. Now such a band can only pull a particle towards a definite position if it is fixed at its other end. It may happen that this other point is also a material particle capable of being moved, and which is connected to other particles in turn—Fig. 30. When therefore any one of the particles is displaced from its position of rest by means of a blow, that portion of the rubber band connecting it to its neighbours is stretched, and there are noticeable at both ends of the rubber forces which tend to draw the rubber and the first particle back to their equilibrium positions; but this same tension at the other end of the rubber band causes the second particle to be displaced from its position of rest. Since, however, the first particle has already moved away from its zero position before the second one has begun to move, this second particle is retarded with respect to the first one. This second particle is attached by means of an exactly similar band to the third, and, as soon as it has left its zero position, a force will be executed upon this third particle which begins to move more slowly than the second one. The process is continued until the whole series of particles is executing a normal harmonic motion, but each one is retarded with respect to its predecessor. The complete movement is referred to as a wave motion.

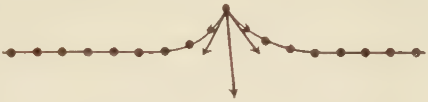


Fig. 30.

If a stone is thrown into a weir, the disturbance is propagated over the surface of the water in the form of rings. The propagation of sound, light, or the signals used in wireless telegraphy and telephony, are the results of such motions. All these different waves vary with respect to one another : not only is the type of particle different but the nature of the restraining forces also varies according to the nature of the particular waves which are being propagated. There are, however, a whole series of laws which are common to all wave motions. We shall now proceed to discuss these laws, and in order to obtain a clear conception of the processes which are at work we shall consider the vibrations of material particles placed at regular intervals along an elastic cord.

The Different Kinds of Waves. It may happen that the particles which are excited by the motion of a material particle surround that particle on all sides and that these, in their turn, excite the particles which surround them. Under these conditions the disturbance will be propagated in all directions, and when the velocity of the disturbance is the same in all directions, all the particles which are being excited at any one instant will lie on the surface of a sphere, at whose centre the source of all the waves is situated. According to the extent to which the disturbance travels forward, the radius of this sphere increases. The phenomena of sound and light are the results of such disturbances.

If, at first, we only consider the particles affected as lying in one plane, the wave will be propagated in the same manner as waves upon the surface of water. The stretched skins of drums vibrate in this manner also.

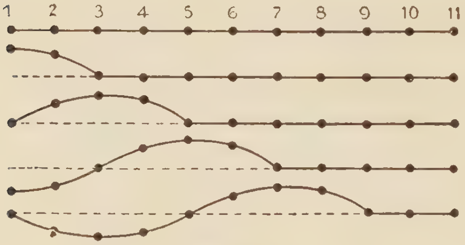


FIG. 31.

Finally there are waves which can only travel in one direction because there are no particles present whereby they can be transmitted in any other direction. Such motions occur in strings or ropes, as in pianos, violins, etc.

Suppose that such a process is represented in Fig. 31. For the sake of greater simplicity we will assume that the masses are concentrated in points, and that they are bound together by pieces of elastic in which the forces arise. (I) Suppose that some external agency causes the first point to execute a normal harmonic motion. After a quarter of a period (II) it has caused the following two masses to move, and after a half period it has returned to its zero position (III). Meanwhile, however, the third point has undergone its maximum displacement and the motion will have extended to the fifth particle. When three quarters of a period have elapsed, the first particle has moved to its extreme position on the other side of the zero line and the disturbance has advanced as far as the seventh particle. When the first point has completed a whole vibration the motion has advanced as far as the ninth particle. Since the particles 8, 7, 6 . . . have always been moving for a longer time than the particle immediately succeeding them, it follows that all the states of motion which have been associated with the first particle in the whole process, are present at some point or other in the medium at the same instant.

It is necessary for us to distinguish quite clearly for all waves, the difference between the motion of each individual point and the propagation of the wave as a whole. For example, the highest points on a circular wave moving on the surface of water always lie upon a circle, the periphery of which is moving outwards. But it is not any one particle which is moving forwards, but a certain condition of maximum displacement which continually advances from one drop of water to another more remote from the centre of the disturbance. If a cork or leaf is thrown upon the surface of the water it is easily seen that the cork is not carried forward by the wave, but merely dances up and down always remaining at the same distance from the centre of the disturbance.

Transverse and Longitudinal Waves. The above example shows very clearly that the two motions have different directions. The cork moves up and down perpendicularly to the surface of the water, whilst the wave itself is transmitted in all directions across the water: because these two directions are mutually at right angles the wave is referred to as a **transverse wave.**

It is not at all necessary for every wave to move in this manner, for the wave may move in the same direction as the particles themselves. When a hammer hits a long rod, the blow being aimed along the length of the rod, the particles are displaced along the axis of the rod and the wave travels in this direction. Such waves are called **longitudinal waves,** and the waves which are propagated through air belong to this category.

In the case of longitudinal waves it is very necessary for us to distinguish between the movements of the particles themselves, and the propagation of the rarefactions and compressions which really constitute the waves; this question is all the more important in this particular type of motion, because both movements occur in the same direction, and are therefore not so readily visualized as with transverse waves. Fig. 32 shows the distribution of the layers in a longitudinal wave.

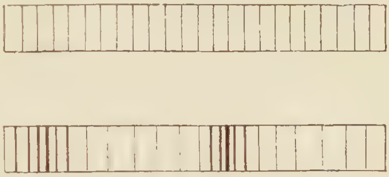


Fig. 32.

In order for a transverse wave to arise in a medium, it is necessary that the motion of one particle in a given direction must entail the displacement from their rest position of particles lying perpendicular to this direction. This only happens with solid bodies. With liquids and gases the particles are more readily shifted, there being no appreciable forces tending to prevent such a disturbance. It therefore follows that transverse waves can only be transmitted through solids.

Longitudinal waves, however, result from the action of a moving particle on other particles in its direction of motion. This happens whenever forces are available which counteract the nearer approach of the particles to one another, *i.e.* wherever bulk elasticity is present. Such is the case with liquids and gases, as well as with solids.

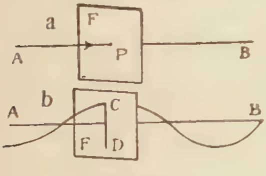


Fig. 33.

Polarization. A very important difference between longitudinal and transverse waves must now be mentioned. Let AB be the direction in which a wave is moving, and consider a plane F which is perpendicular to this direction. If the wave is longitudinal, Fig. 33 (a), the motion of the particles is always perpendicular to this plane. There is no indication of this wave in the plane except at the point P through which it passes. If the wave is transverse, however, Fig. 33 (b), the motion of the individual particles takes place in this plane; a vibrating particle leaves a trace in this plane which marks the path which it has traversed, and this direction, CD, which is perpendicular to the line of propagation, is now distinguished from all other directions in this plane. The wave is said to be **polarized.** The phenomenon is of great importance when we have to decide the nature of an unknown wave motion; if the phenomenon of polarization can be demonstrated, then the wave is certainly a transverse one.

A vibrating material particle in an extended medium is so intimately connected with all the neighbouring points, that its energy is always transferred straightway to its neighbours; it therefore does not execute an independent vibration, and an extended medium has no period proper to itself. Any particle can be made to vibrate with any particular period, and the disturbance will be propagated throughout the whole medium without the period being affected. High and low notes are transmitted equally well in free air.

Under such conditions the whole of the existing energy is transmitted to the next section of the wave, *i.e.* no portion of the energy is propagated in preference to any other portion. If this had not been so it would have been impossible to receive a succession of notes in their true order, and intelligible communication of external events by means of sound waves would have been impossible.

The Fundamental Laws concerning Waves. We will assume that a vibrating point is executing a normal harmonic motion the period of which is T. If the period is short, one prefers to speak of the **frequency,** ν, *i.e.* the number of oscillations per second. Hence

$$\nu\, T = 1 \qquad \text{. . . . . (42)}$$

When the first particle has made one complete oscillation the disturbance has arrived at the ninth particle (*cf.* Fig. 31, p. 74) and all the states of motion through which the first particle has passed during the time T exist somewhere amongst the first nine particles. This distance is called a **wave-length. The wave-length is therefore that distance through which the disturbance is propagated in the time T.**

The velocity with which a wave is propagated through a medium tells us at once how far the wave has travelled in unit time. If we denote it by c, then

$$c = \nu\lambda = \frac{\lambda}{T} \qquad \text{. . . . . (43)}$$

If the first particle executes a vibratory motion represented by the equation $y = a \sin \frac{2\pi}{T} t$, the particles which lie to the right of this particle undergo displacement represented by this same equation, but at a later time. The difference between the times is of course equal to the time required for the wave to travel from the first particle to the other. If the distance apart is x cm., and the velocity of propagation, *i.e.* the distance traversed in unit time, c, then this difference is x/c sec. The time t which has elapsed since the first particle began to move, must therefore be diminished by this same amount. The displacement of any point at distance x from O is therefore expressed by

$$y = a \sin \frac{2\pi}{T}\left(t - \frac{x}{c}\right) = a \sin 2\pi\left(\frac{t}{T} - \frac{x}{\lambda}\right) \qquad \text{. . (44)}$$

If, in this equation, the value for t is fixed, and x assumes all values between O and λ, then y represents the different states of motion for the different points in the wave at this instant: in fact it gives us an instantaneous picture of the wave.

If x is fixed whilst t assumes all values from 0 to T, then y denotes the variations in the displacement which occur at any particular point throughout one period of the oscillation: it gives us a kinematic picture of the motion at one point.

Compound Wave Motions. Since the principle of independent motions is quite general, it follows that when two wave motions occur at any one place, the resultant effect will be due to both motions. The final wave form will depend upon the amplitude, wave-length, the difference in phase, and the direction of propagation, which characterize the waves arriving at the point under consideration. It follows, therefore, that the resultant effect may be very complex. **Fourier** (1768–1830) has shown that the oscillations of any string can be resolved into a series of component oscillations the periods of which bear a simple relationship to one another.

A few of the simplest and more important examples of the superposition of vibratory movements will now be briefly mentioned. In all cases it will be assumed that the velocity of propagation is the same, and further, the waves will be treated first as waves of the same wave-length travelling in the same direction. Let us consider the vibration of a long string which has been stretched tightly, Fig. 34. The displacement from the position of rest will, at all places, be equal to the sum of the constituent displacements. In every case will appear a new wave, having the same wave-length and the same velocity of propagation.

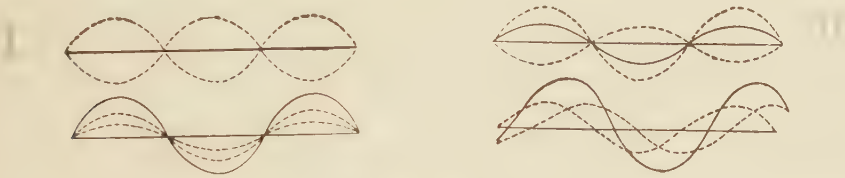


Fig. 34.

In I the amplitudes of the constituent waves are equal, but the form of the one wave is a mirror reflexion of the other, *i.e.* there is a phase difference of half a wave-length between the two waves. Since the displacements are everywhere equal, but opposite, the string remains at rest. This is a very important result, for it shows that two equal waves, having a phase difference of half a wave length, completely neutralize one another.

The second example considered only differs from the first in that the two component waves have different amplitudes; they do not efface one another completely, but the amplitude of the more intense wave is reduced.

In III the phase difference is zero but the amplitudes of the constituent waves differ. The amplitude is simply increased.

When there is a difference of phase between the two waves, as in IV, then a new wave form of the same wave length is produced. This differs in phase and amplitude from the original waves, and its particular form depends upon their amplitudes and phase differences.

Stationary Waves. A very important type of wave motion occurs when two equal waves are transmitted in opposite directions through a medium. Let us assume that the wave which is represented by dots, Fig. 35 (I) is travelling from right to left along a string, whilst the dashes represent the wave in the reverse direction. Let P be the place where the two waves first encounter one another (II). In order to determine the state of motion at P, at any later time, it is only necessary for us to observe that during this time both waves have advanced by equal amounts, the one towards the left, the other towards the right. For example, after a time equal to one quarter the periodic time, the crest of each wave has just arrived at P. The effect at this point is now a maximum. Later on, the displacement becomes less, because the two component waves begin to decrease together, so that after a time $T/2$ the string has returned to its position of rest. After a time $\frac{3}{4}T$ the displacement is a maximum on the other side of the zero position; the string then begins its return journey. The resultant effect is that P executes a normal harmonic motion whose amplitude is twice that of the constituent waves.

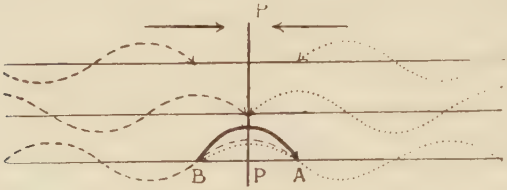


Fig. 35.

Now let us consider the state of things at A, a point situated at a distance $\lambda/4$ from P. Interference begins to take place at A, $T/4$ sec. after the waves first met at P. In the meantime the wave which is proceeding from the right has moved a distance $\lambda/4$. Fig. 35 (III) shows us the condition of affairs at this instant. We see that the point A remains at rest when interference first occurs. When one desires to know the state of affairs at some later time it is only necessary to find the points in the two waves which have arrived at A; these will be such that they were at an equal distance from A, but on opposite sides, when the waves first met. It is always found that the displacements at these points are equal and opposite, so that the point A is always at rest, *i.e.* at this position the two waves efface one another continuously. The same state of affairs is found at B, a point distant $\lambda/4$ towards the left of P. The distance between these two points A and B, which are always at rest, is half a wave-length. Towards the right and left of these points the string is moving to and fro with varying displacements; at the middle the displacement is a maximum, viz. twice that of the individual waves. The stationary points are called **nodes**, and they are found at distances apart equal to half a wave-length. If the length of the string is an integral multiple of half wave-lengths, then this number of half waves can be observed in the string, each one terminating in a node. Since the waves do not present the impression of waves moving along the string, they are called **stationary-waves.** They are always formed when the medium in which waves travel has a boundary. It is only the great natural expanses such as the atmosphere, ocean, or interstellar spaces, which can be considered as media having no boundary.

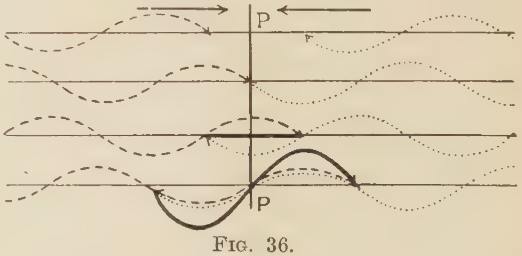


Fig. 36.

In order to supplement what we have already discussed, let us consider two waves travelling in opposite directions and with a phase difference equal to half a wave-length, Fig. 36 (I). The waves first encounter one another at P, Fig. 36 (II). After a time $T/4$, Fig. 36 (III), the amplitudes of the waves arriving at P are equal and opposite, a condition which is permanently maintained at this point. After a still further increase in the $T/4$ we have a state of motion represented by IV. The interference which takes place at all other points is easily explained; it is found that when two waves, differing in phase by $\lambda/4$, meet, the point where they first encounter becomes a node, whereas it is a loop or antinode when the constituent waves are equal in phase.

2. THE TRANSMISSION OF WAVES

Free Transmission in a Homogeneous Medium. The meaning of the term **velocity of propagation** is very important in connexion with such a motion. **Newton** was acquainted with the fact that the velocity c of a wave, in a medium having an elastic coefficient e and a density d, *i.e.* the mass per c.cm., is given by the formula $c = \sqrt{\frac{e}{d}}$, *i.e.* the square root of the quotient formed by dividing the elasticity by the density. In order to prove this equation in a particular case, let us assume that the medium extends to infinity but along a straight line only. By making the medium infinite we do not have to consider the complication arising from possible reflexions which may occur at the boundary.

Suppose that Fig. 37 represents a rubber rod. In such a medium opposing forces arise when one endeavours to press together the individual sections into which it might be supposed that such a medium has been divided. For the sake of clearness we will assume that these sections have been extended considerably, so that elastic forces must have been called into play amongst them. Let us assume that a compressional wave has been set up by applying pressure with a piston at the end A. Suppose that the piston has moved forward with a velocity v during a time t. The first layer will commence to move with velocity v and in endeavouring to cause the second one to move elastic forces will arise. These act upon the second layer and cause it to move; but they also react upon the first section and would bring it to rest, were it not for the fact that the piston is exerting an equal pressure from the other side. The first section is therefore being compressed by equal forces so that it is no longer accelerated but continues to move forward with velocity v. The second layer is set in motion and, in endeavouring to make the third one move, produces an opposing force which reacts upon itself. As soon as it has acquired a velocity v the opposing force is equal to the one which is urging it forward, so that, like the first section, it moves forward with a velocity v. The process is continued. How quickly a motion is imparted to the succeeding particles depends, firstly, on the magnitude of the elastic forces

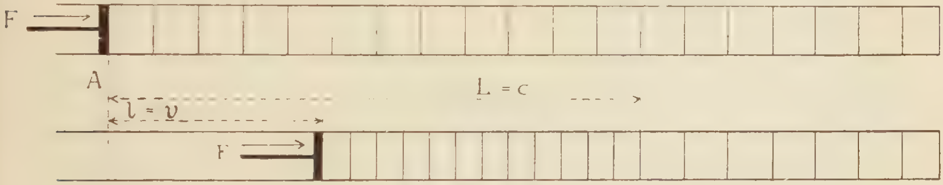


FIG. 37.

which have been produced, and secondly on the inertia, or mass, of the individual particles. If the velocity of propagation is c, then in one second the disturbance will have travelled a distance $L = c$ along the cylinder. In this length of the cylinder the layers have been somewhat compressed and all of them are moving with velocity c. Meanwhile the piston has moved a distance $l = v$, and if F is the force which has been applied to it, the work done is $F.l = F.v$. This work has been expended, partly in compressing the elements of the elastic together, and partly in accelerating them. Owing to the first part, the elastic has been reduced in length from c by an amount v. According to equation (38) developed on page 67, this corresponds to an increase in potential energy equal to $Fv/2$. From this it follows that exactly one-half of the external work done has been spent in overcoming the elastic forces ; the second half served to accelerate the particles. We then have

$$\tfrac{1}{2}Fv = \tfrac{1}{2}mv^2, \text{ or } F = mv \quad . \quad . \quad . \quad . \quad (45)$$

It is now only necessary to insert the value for F according to (36), remembering that $l/L = v/c$. We therefore have

$$F = \frac{eav}{c}.$$

Now the mass of the column of length c is $m = cad$, where a is the area of cross section and d is the density. Inserting these values of F and m in (45), we have

$$\frac{eav}{c} = cadv,$$

whence

$$c = \sqrt{\frac{e}{d}} \quad . \quad . \quad . \quad . \quad . \quad (46)$$

If this expression is examined it will be observed that the velocity v of the piston is not contained therein. This very important result shows that the velocity with which a disturbance is transmitted through an elastic medium is independent of the manner in which the disturbance arises. The factors which are important are e and d, and these are properties of the medium itself. It therefore follows that by measuring the speed of such a velocity, no conclusion can be drawn concerning the nature of the disturbing agent, but only concerning the nature of the medium. If two unknown wave motions have different velocities, then, from the very beginning, one can be quite certain that two essentially different phenomena must be occurring in different media. But if the two velocities are equal it does not follow at once that the two motions are occurring in the same media, for it is just possible that two different media may be capable of transmitting the disturbances with equal velocities, *i.e.* the ratio $\frac{e}{d}$ would be the same in each case. But such 'chance' coincidences are of very rare occurrence. When **Maxwell** learnt that **Weber** and **Kohlrausch,** working in Göttingen, had found a value for the ratio of the electrostatic and electromagnetic units, which was equal to the velocity of light, he concluded at once that the transmission of light must be electromagnetic in nature. Maxwell thereupon developed his electromagnetic theory of light, which is one of the greatest advances that have ever been made in science ; he proved that light and electromagnetic waves were essentially the same.

The Transmission of Waves in a Three-Dimensional Medium. Huygens' Principle. Let us consider a material particle placed in a medium which has no boundaries anywhere. If such a point is displaced from its position of rest, the particles in its vicinity will be affected. Let us assume that, at some instant, the disturbance has reached the surface of a sphere whose radius is r. Then all points inside this space are each equivalent to the first, for each one may be regarded as the centre of a disturbance. It might be imagined that an absolutely inextricable array of spherical waves would arise, which, for example, would not be at all suited for the transmission of speech. The Dutch physicist, **Christian Huygens** (1629–1695), a contemporary of Newton, first solved this problem, and the French physicist, **Fresnel**, extended and completed his work.

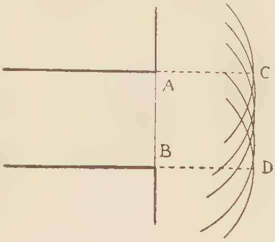


Fig. 38.

Let AB, Fig. 38, be a section of a portion of a plane wave. This line really denotes the position of a whole series of points which, at the time chosen, are just about to be disturbed by the wave. This line, which is the trace of a surface, is called a wave-front. According to Huygens, every point in this line is the centre of a new spherical wave, but each of these waves alone is not sufficient, on its own accord, to stimulate one of our senses. This will only happen when these disturbances all exert their influence simultaneously at some position in front of AB. This can only occur at CD, to which surface the elementary disturbances sent out from AB all arrive together. In this manner Huygens was able to explain the rectilinear propagation of light, its reflexion and refraction, together with other light phenomena.

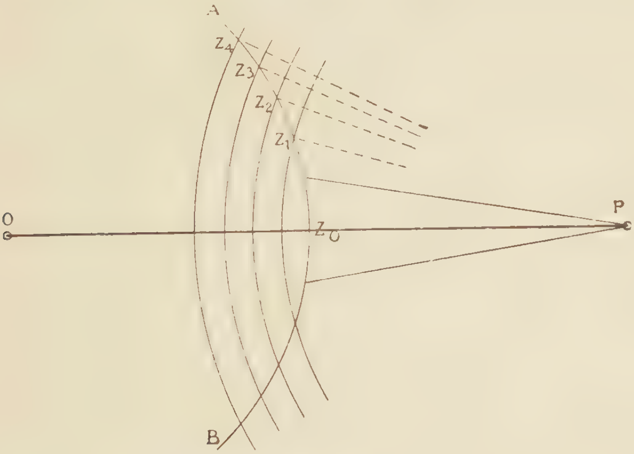


Fig. 39.

However important this first advance in a hitherto uninvestigated field was, it only permitted a satisfactory explanation to be given, when Fresnel observed and took account of the interference of waves, and used this phenomenon in his explanation. Fresnel pointed out that if one wishes to ascertain the state of affairs at any point P, Fig. 39, then it is necessary to contemplate the influence of all points in the wave upon this particular point P. But, if different disturbances are arriving at a point, the resultant effect will depend upon the intensity of the individual wavelets and their phase differences. For example, if two equal waves with a phase difference equal to half a wave-length meet at a point, they will not strengthen one another, but on the contrary will mutually destroy each other completely.

If, therefore, AB represents the front of a spherical wave whose centre of disturbance is at O, then all points on AB are invariably in the same phase. The wavelets which proceed from all the points on AB will, in general, be in phase at points whose distances from the surface AB are either equal, or else differ by any number of complete wave-lengths. At those points, however, where the difference in path between two disturbances amounts to a half wave-length, or an uneven number of half wave-lengths, the two disturbances efface one another completely. Fresnel divided the whole surface AB into zones, the limits of which were so chosen that a difference in path of half a wave-length existed between one zone and its immediate neighbour. Let PZ_0, PZ_1, PZ_2, PZ_3 . . . be distances such that each one is half a wave-length longer than its predecessor. Then, for example, a disturbance from Z_3 must travel a distance longer by half a wave-length than does a wave from Z_2, in order to arrive at P. When therefore the wave from Z_3 arrives at P, the wave from Z_2 has already passed through this position and is a distance $\lambda/2$ ahead. But at the point P itself a wave from Z_2 is still arriving, and this, differing in phase by half a wave-length from the disturbance which has just arrived from Z_3, causes the two wavelets to destroy one another. The whole surface is divided into a series of circular zones; the innermost zone is part of a spherical surface and its centre lies upon the line OP. The following zones are all strips which surround the first one, but on account of the curvature of the spherical surface the breadth of these zones decreases as one proceeds outwards. Fresnel has shown, by paying attention to this diminution in breadth and the phase differences which exist, that the effect of any one strip is exactly destroyed by half the collective effect of the disturbances from its two immediate neighbours. There is one exception to this rule, namely, that one half of the effect of the central zone is never annihilated. If, therefore, S_0, S_1, S_2, S_3 . . . denote the effects of the different zones, then the total effect is, according to Fresnel's assumption,

$$S_0 + S_1 + S_2 + S_3 + \ldots = \frac{S_0}{2} + \left(\frac{S_0}{2} + S_1 + \frac{S_2}{2}\right) + \left(\frac{S_2}{2} + S_3 + \frac{S_4}{4}\right) + \ldots$$

The terms in brackets represent the effects arising from one zone, and the half-effects arising from its two neighbours, so that the terms in every bracket must, as Fresnel showed, vanish.

From this it follows that the total effect of all the zones is

$$S_0 + S_1 + S_2 + S_3 + \ldots = \frac{S_0}{2} \quad . \quad . \quad . \quad (47)$$

This equation shows that the total effect is equal to one half of that arising from the central zone.

The above analysis accounts at once for the apparent rectilinear propagation of a disturbance in a medium which extends to infinity on all sides. Actually the disturbance is transmitted in all directions outwards in the form of spherical waves, but the effects at points other than in directions deviating from a straight line balance out. The following phenomenon yields a striking proof of the correctness of these considerations.

If the resultant effect of all the zones is altered by making the zones 2, 4, 6 . . . opaque to the particular type of radiation we are considering (*e.g.* the light is transmitted through a glass plate in which the zones are alternately clear and black), then the disturbances which arrive at *P* are always in phase ,for they differ by a whole number of wave-lengths. The point *P* will therefore be the focus for the zone plate for light of the particular colour for which the plate was designed. These facts will be verified further when we come to discuss the subject of light more fully.

3. THE DISTURBED PROPAGATION OF WAVES

What Happens at the Boundary of a Medium ? Most of the media in which it is possible for us to transmit a disturbance are only of finite extent, so that, after a short time, the waves must arrive at the boundary of the medium in which they have travelled. It is impossible for the waves to vanish suddenly, for that would imply a violation of the law known as the Principle of the Conservation of Energy. What will happen ?

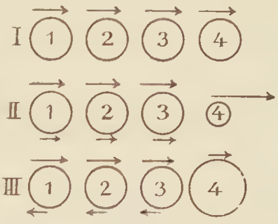


FIG. 40.

In order to understand the problem let us remind ourselves of the law of elastic impact. The case of a sphere which was at rest before impact occurred is more than similar to the problem which we are now discussing. For, at the instant of impact, the two spheres exert an influence upon one another in virtue of the elastic forces which arise, in exactly the same way as, in the case of wave transmission, the individual particles exert an influence upon their immediate neighbours. Even in this problem the particles are at rest until they become stimulated by their neighbours. If the medium is homogeneous, then the problem resembles that of two equal colliding spheres. But if the medium has different properties at any point, *e.g.* its elasticity or density may vary, then the problem is similar to that in which impact occurs between two spheres, one of which is less massive than the other. If the two spheres are equal and moving in opposite directions, the colliding sphere is brought to rest. If the sphere which is hit is not so heavy as the other, then this latter retains part of its velocity and proceeds in its original direction but with diminished velocity. If the sphere which is hit is heavier than the colliding sphere, then this latter sphere retraces its original path. The two extreme cases of such a collision are as follows : if the sphere which receives the blow is infinitely massive, as in the case when impact occurs upon a solid wall, then the colliding sphere returns with the velocity it had on arriving at the wall ; if the sphere which receives the blow is very small (in the limit its mass is zero), then the colliding sphere continues its journey with a velocity which has only been slightly diminished (or not at all).

In a homogeneous medium the first particle collides with the second, the second with the third. . . . On each occasion the particle which is hit has a mass equal to that of the colliding particle, but it was at rest prior to the collision. [This is represented in Fig. 40 (I), by arrows drawn above the spheres.] The colliding particle, however, gives up its energy to the particle which it encounters and is consequently brought to rest, but the wave continues in the same direction as before. If, let us say, the fourth particle, however, has a smaller mass, then the colliding sphere is not brought to rest but still moves along towards the right [arrows below the small circles in Fig. 40 (II)] and attracts the second particle, which had been reduced to rest ; hence a new wave motion arises which is propagated towards the right. In consequence of its action upon the third particle, this latter is brought to rest. The second particle then affects the first particle : in short, a reflected wave travelling backwards is formed, which travels towards the right whereby all the particles are deflected to the right as at first, but the excitation itself travels to the left.

If, however, there is a heavier particle at some position in the medium, Fig. 40 (III), [or if the impact occurs upon an 'infinitely massive' wall] the third particle begins to retrace its path after impact has occurred, and collides anew with the second particle, which was already at rest. Then the second particle collides with the first, so that all the particles now move backwards and participate in a wave which travels in a direction opposite to that of the original one.

We have thus shown that the boundary of a medium, *i.e.* a place where the medium ceases entirely, or where another medium begins, is the seat of a new wave motion. If the new medium is more dense than the first. then the particles move in a direction which is opposite to that in which they originally moved ; if the medium is less dense, then the particles move in the same direction as the original wave did. The term 'more dense' is used to imply a medium in which the density is greater, and, in virtue of this, the velocity of propagation is diminished. A diminution in velocity can also take place if the elasticity of the medium is diminished whilst the actual density remains unchanged. In such cases, however, it is customary, for the sake of clearness, to speak of the medium as being 'more dense'.

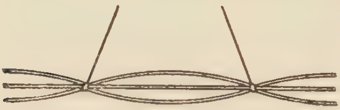


Fig. 41.

In the study of sound waves, the wave which is reflected from a wall or boundary is called an echo. If it is necessary to specify the time which elapses before a sound wave returns to its source, it is only necessary to divide the whole distance from the source to the reflecting wall and back by the velocity of sound. When it is necessary to specify the phase of the wave at any point on the return journey, then two cases must be distinguished. If the reflexion occurs at a denser medium, then the distance must be increased by half a wave-length ; it is not necessary to do this when the reflexion occurs at a less dense medium.

When an uninterrupted train of waves is being emitted from some source and reflexion occurs at the boundary of the medium, then the reflected waves must interfere with those which are still being emitted. For example, let us consider the vibrations of a stretched string, which will behave in the manner already indicated. There will be two waves of equal wave-length, but travelling in opposite directions. The point at which the waves first meet will be the end-point of the medium, *i.e.* the point from which the return waves start. If this end-point is at the boundary of a less dense medium, or a medium with zero density, as, for example, happens at the free end of a steel rod, then points in the return wave are moving towards the same side as points in the arriving wave. Such a motion is shown in Fig. 41. The point at the boundary is an anti-node, or loop. If, however, the reflexion occurs at a more dense medium, as, for example, occurs when a disturbance reaches the bridge which supports a violin string, then points in the return wave move in a direction which is opposite to that of points on the incident wave. The point where the waves first meet, viz. the end of the medium, is therefore a node (Fig. 42).

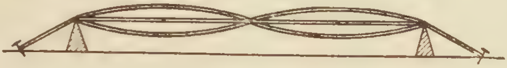


Fig. 42.

The Reflexion of Spherical Waves. Let us assume that the point O, Fig. 43, is the source of spherical waves. At distance OB let the first medium be bounded along MN by a medium of infinitely great density. The first point of contact between the spherical waves and the boundary is at B; the reflected wave therefore begins at B, whilst other portions of the spherical wave are still being propagated as before. Finally the points M and N are reached. If the boundary had not been present, then the wave front would have been at MCN, and that part of the disturbance which would have been below the plane MN, has now been formed above MN, and represents the wave front of the return wave. The centre of this spherical segment is at O_1, where $O_1B = OB$. The reflected wave will thereby travel through the original medium as if it had been a wave originating in O_1, a point which is as far behind the surface MN as O is in front of it.

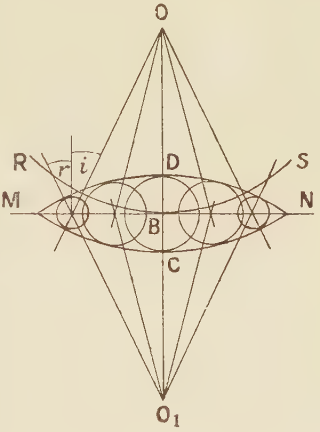


Fig. 43.

If one wishes to indicate, let us say, the path taken by a narrow cone of rays originating at L, Fig. 44, and which enter the eye of an observer after reflexion at MN, then the straight line LA is drawn perpendicular to MN and produced to L_1, so that $AL_1 = LA$. Then a cone whose vertex is at L_1 and whose base is the pupil of the eye, indicates the pencil of rays which enter the observer's eye. The angle included between the incident ray (*i.e.* a very narrow cone of rays) and the normal to the surface at the point of incidence is called the angle of incidence. The angle between the normal and the reflected ray is the angle of reflexion, and the fact that the triangles LAB and L_1AB are congruent shows that the angle of incidence is equal to the angle of reflexion.

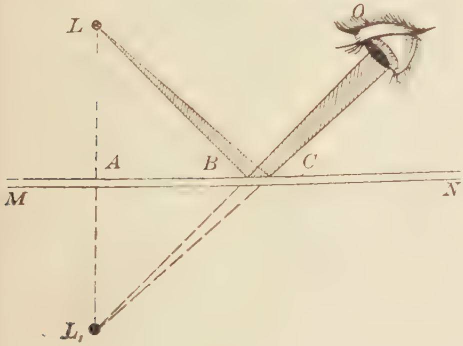


Fig. 44.

The Refraction of Waves. Let us assume that a wave is incident at an angle i, upon the boundary, MN, of a new medium which is more dense than the first, but that the velocity of transmission in the second medium is less than that in the first. Suppose that the boundary is a plane surface; the problem is to discover the direction in which the wave is transmitted in the second medium.

It will be assumed that the source of the waves is at so great a distance from *MN* that the wave front can be considered as plane. The light waves which we receive from the sun can be regarded as such. Let *AB*, a straight line perpendicular to the direction of propagation, be the wave front at one instant; a little later it is represented by *CD*, where *C* is the first point of incidence. Now although the plane wave has still to move in the original medium, a spherical wave, whose source is *C*, has begun to travel in the second medium, but with diminished velocity. If the ratio of the velocities in the two media is $c : c' = 3 : 2$, say, then must $DE : CF = 3 : 2$. The spherical wave originating at *C* travels a distance, $CF = \frac{2}{3}DE$, in the second medium before the disturbance in the first medium has reached *E*. But the point *E* is also a point in the new medium. In the same way other points on the refracted wave front can be determined; the tangent drawn from *E* to the circle whose radius is *CF* represents the wave front in the second medium. The refracted disturbance is propagated perpendicular to this direction.

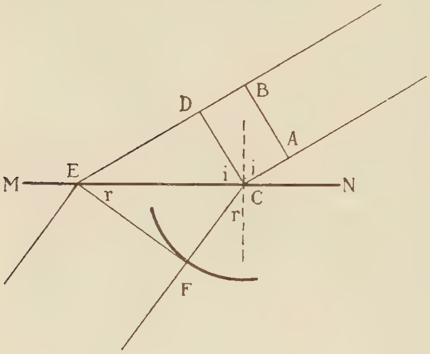


Fig. 45.

From the geometrical relationships which exist in Fig. 45, the following connexion between the angles of incidence and refraction is obtained. If at *C* a normal to the surface *MN* is drawn, then *i* is the angle of incidence, and *r* the angle of refraction.

Now $$i = D\widehat{C}E, \text{ and } r = C\widehat{E}F$$

But $$\sin D\widehat{C}E = \frac{DE}{EC} \text{ and } \sin C\widehat{E}F = \frac{CF}{EC}, \text{ so that } \frac{\sin i}{\sin r} = \frac{DE}{CF} = \frac{c}{c'}.$$

Thus the ratio of the sines of the angles of incidence and reflexion is constant (**Snell's law,** 1626) and this ratio is equal to the ratio of the velocities of propagation in the two media.

This ratio of the velocities is the refractive index of the medium so that the law of refraction may be expressed by

$$\sin i = \mu \sin r \quad . \quad . \quad . \quad (48)$$

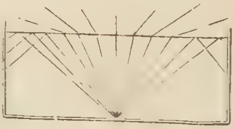


Fig. 46.

When waves from a source in a dense medium are incident upon the boundary of a less dense medium, Fig. 46, the direction of the refracted waves is again determined by the above formula. When the angle of incidence is so large that $\mu \sin r$ is greater than unity, then refraction can no longer take place. The greatest value for *r*, which is now the angle of incidence really, is given by $\mu \sin r = 1$, or $\sin r = 1/\mu$. The corresponding angle of refraction (now *i*) is expressed by $i = 90°$ and the wave is transmitted along the boundary of the two media. If the angle *r* is greater than that defined by this limiting condition, then no refraction can take place and the wave is totally reflected at the boundary. This phenomenon, which is of considerable importance in the study of light, is referred to as total reflexion.

Waves and Rays. It must be admitted that, in all these discussions, a certain contradiction to our customary mode of thought and expression presents itself. We consider light as a process of radiation. Rays are emitted from the sun and these are focussed by lenses so that they form a bundle or pencil of rays. The rays which were parallel to one another when they encountered the lens now become either convergent or divergent. According to the manner in which we have presented the subject there are no rays, but only wave surfaces which may be plane, convex or concave. It is only in an extended wave front that any appreciable amount of energy is to be found. When it is necessary to consider the real state of affairs, it is better to use the idea of wave surfaces. There are no rays of light, in so far as we understand by that term formations which extend in one direction only, *i.e.* along the beam and not perpendicular to it. In particular, mention may be made of interference phenomena, which cannot be interpreted in terms of rays, but which are easily explained in terms of the wave theory. Nevertheless the idea of a ray of light, heat, or electrical energy, can be used when it is only a question of a convenient expression rather than a description of the processes, more especially since, according to the wave theory, it is only the energy associated with a very limited portion of the wave surface that travels in a straight line from the source to the eye of the observer.

In the sequel we shall often speak of rays but we shall remember that it is only a short and convenient mode of expression, rather than a statement of the real facts.

CHAPTER VII

ACOUSTICS

1. Sound and Music

THE ear is the organ of the body which enables us to become cognizant of one special group of happenings which take place in the world around us. Simultaneously with the sensation produced in its ear, a child notices the movements of its mother's lips and so, quite unconsciously, learns to interpret these sensations in its ear as representing happenings in the external world. Very soon we learn to distinguish small differences in our observations on sound, and know at once whether a noise is caused by something close at hand, or whether its origin is more remote ; we are also able to say whether its source is to our right or left. We are able to distinguish the voices of our acquaintances from those of other people. We also find that we ourselves can produce sound by means of our vocal organs, and from the behaviour of other men, and even animals, we deduce that their ears enable them to make similar observations.

Physics endeavours to discover what external processes are the cause of these perceptions. These processes are called sound. We shall now discuss what external events are able to produce a sensation in an ear, then we shall say how sound is propagated through the intervening medium, and finally we shall discuss the origin of the sound waves which proceed from the source to us.

In our observations on sound we distinguish three different qualities which are called the pitch, intensity, and quality ; each of these will now be briefly discussed.

The **pitch** is the most important of these because it is that property which enables us most frequently and efficiently to recognize a particular sound. It was recognized quite early in the history of this science that our ear perceives a sound when a rapid and definite succession of disturbances reach it. Generally speaking the ear is stimulated by the surrounding air which has acquired a vibratory motion. This motion is transmitted to the ear drum which begins to vibrate with the same frequency as that of the sound which it receives. But this is not necessarily always so, for if a tuning fork is struck, and the handle held against our head or teeth, the pitch is very easily and definitely observed. Now it is easy to observe that a high-pitched tone is produced when a rapid succession of disturbances reach the ear, and that a deep tone is caused by disturbances which follow one another more slowly. This fact was known to **Pythagoras** (580–500 B.C.), and **Aristotle** knew that a stretched wire gave a note which was an octave higher than its fundamental when its length had been diminished by one half. But the exact mathematical relationship between frequency and pitch has only been known since the Middle Ages. The Franciscan monk **Fr. Mersenne** (1588–1648) compared the pitch of the note produced by wires of different lengths, but of the same material, and under the same tension, but more exact measurements were first made possible when the siren was invented. **Robison** (1799) used a siren in the form of a toothed wheel, whilst **Cagniard de la Tour** used a wheel in which a number of equidistant holes, near the periphery, had been bored.

Musical Intervals. Musical Notation. Let us suppose that a musical note is produced when the number of vibrations is m per second. We will denote this note by the letter C. If another note is produced when n $(n > m)$ vibrations occur per second, then the **ratio** $n : m$ is called the **interval** between the two notes. Now it does not follow that any two notes, when sounded together, will produce a pleasant sensation, *i.e.* they may not be **concordant**. It is found, however, that when $n : m$ is equal to $2 : 1$, $4 : 1$, $8 : 1$, etc., the notes are regarded as being the same from a musician's point of view. If the note whose frequency is $2m$ is called c, then the interval between the two notes is an **octave.** The note corresponding to a frequency $4m$ is called c^1, etc. For other notes, whose frequencies are p, q, r respectively, it is found that the musical effect produced, when all these are sounded simultaneously, will be concordant if $p : q : r = 4 : 5 : 6$.

The notes which comprise an octave are represented by the letters C, D, E, F, G, A, B, small letters with dashes, or suffixes, denoted the octaves above and below. The whole is referred to as a **diatonic scale.** If m is the frequency of the note C, then those of the others are as given below.

C	D	E	F	G	A	B	c
m	$\frac{9}{8}$m	$\frac{5}{4}$m	$\frac{4}{3}$m	$\frac{3}{2}$m	$\frac{5}{3}$m	$\frac{15}{8}$m	2m

The intervals between the notes are respectively :

C to D,	D to E,	E to F,	F to G,	G to A,	A to B,	B to c
$\frac{9}{8}$	$\frac{10}{9}$	$\frac{16}{15}$	$\frac{9}{8}$	$\frac{10}{9}$	$\frac{9}{8}$	$\frac{16}{15}$.

It will be observed that the numerical relations between the notes are as follows :

C : E : G = 4 : 5 : 6
G : B : d = 4 : 5 : 6
F : A : c = 4 : 5 : 6

When these particular sets of notes, the so-called **harmonic triads,** are sounded together, the effect is pleasant, *i.e.* they are concordant. When such notes, together with a fourth which is an octave above the lowest note in the triad, are sounded, they constitute a **major chord.** The above triads are known as the tonic, dominant, and subdominant triads respectively.

It will also be noticed that the intervals are either $\frac{9}{8}$, $\frac{10}{9}$, or $\frac{16}{15}$. The first two are called a **tone,** the last a **semi-tone.** The two tones are not absolutely equal to each other but differ by an interval called a **comma** : numerically this is equal to $\frac{81}{80}$.

For musical purposes it is convenient to introduce more notes into the diatonic scale. This is done by raising or lowering the notes by an interval equal to $\frac{25}{24}$. Thus C becomes C♯ (C 'sharp') when the pitch is raised by this amount; similarly when, say, *D* is lowered by the same interval we have D♭ (D 'flat').

Musical Temperament. The number of musical notes becomes somewhat unwieldy when the scales are constructed on the above lines. [It will be remembered that any note can be selected as the keynote.] In music the scales are slightly adjusted so that more or less dissonance is introduced, *i.e.* the scale is **tempered.** There are several musical scales in which the dissonance has been reduced to a minimum. The scale of **equal temperament** is generally used. The octaves remain pure, *i.e.* the interval is $\frac{1}{4}$, $\frac{1}{2}$, 2, 4, etc., but eleven notes are introduced between them, each interval being equal to $2^{\frac{1}{12}}$, *i.e.* 1·0595. The twelve notes so formed constitute the **chromatic scale.** In it the major triads have intervals 1 : 1·2599 : 1·4983, instead of 4 : 5 : 6, *i.e.* 1 : 1·25 : 1·50.

The absolute numbers of vibration in notes which we use in music lie between 16 and 16,000 per second; *i.e.* there are about ten octaves. The lowest bass note which can be produced by a human voice has a frequency of about 64 whilst at about four octaves higher a soprano voice may produce a note which attains a frequency of about 1024.

Absolute Frequencies. To-day we are no longer able to doubt the fact that we always become aware of a definite tone when the number of disturbances which reach our ear per second has attained a definite and constant value. It is quite immaterial how these disturbances have been produced; it matters not whether they have been caused by the best musical instruments, or whether they have been produced by some coarse mechanical contrivance; the only necessary condition is that the number of disturbances per unit time shall be definite and constant. Let us cite a few examples.

1. **Savart's Wheel or Siren.** Suppose a piece of cardboard is held against the teeth of a rotating wheel (Fig. 47). Every time a tooth strikes the cardboard the latter is displaced, and so executes an impulsive vibratory motion in consequence of the successive regular impacts. The card causes the air to vibrate with the same frequency which an ear then receives. In the second form of **siren** (Fig. 48) a disc is made to rotate about its axis. Near to the periphery of this disc there is situated a number of equally spaced holes. From a fixed jet a constant blast of air impinges upon the revolving disc. Every time one of the holes comes before the

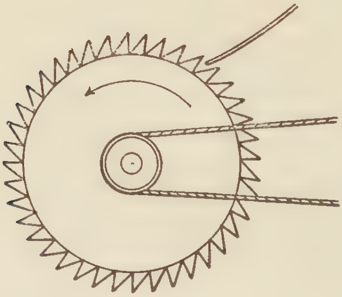


Fig. 47.

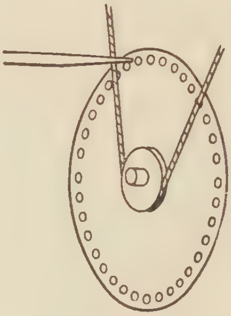


Fig. 48.

jet, the blast of air is able to exert an influence upon the air which lies on the more remote side of the disc. Immediately the hole has passed by, the stream of air is interrupted. There is no instrument present concerning which it could be said that it was vibrating. The sound waves are produced by successive puffs directly in the air itself and one hears a note of the same pitch as if one were listening to any instrument having the same frequency.

2. In the **Trevelyan rocker** (Fig. 49) a hot metal is placed upon a cold support. Every time any part of the support comes into contact with the heated body, it too becomes hot. A slight hump is raised on the cooler body which causes the hot body to tilt; this causes the two bodies to touch elsewhere when the process is repeated. The hot body therefore continues to rock to and fro, emitting a humming sound, until it has cooled to such a temperature that the transmission of heat to the support is no longer sufficient to maintain the motion. With this very crude mechanical contrivance quite a definite and clear note can be produced, the pitch of which depends upon the frequency of the rocking motion.

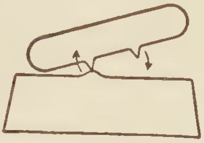


FIG. 49.

3. When gramophone records are being manufactured, the sound from an instrument such as a flute, let us say, is allowed to fall upon a light membrane to which a needle has been fixed. The needle traverses a rotating disc in which it cuts small furrows. In the inverse process, when the plate is being used to reproduce the sound which it has recorded, these wave-like furrows act upon a needle; the needle moves a membrane, which sets the air around it in motion, and finally the air transmits the note to our ear. The more quickly the record is allowed to rotate the higher the pitch; it is therefore possible by altering the speed of the record to convert a tenor into a bass, or even into a soprano.

4. **The Doppler Effect.** A phenomenon which is still more striking is that known as the Doppler effect. Let us assume that we have at our disposal a source of sound, which is continuously giving a note of definite pitch, and let three observers determine this pitch. All three are provided with exactly similar tuning forks; the first observer is travelling upon a ship which is moving down the river, the second travels on a ship which is proceeding up the river with the same velocity as the first ship, whilst the third observer is stationed upon a bridge in line with the ships. The note which is under examination is produced by the siren of a ship which is stationary at some distant point down the river. The frequency of this note is known to be ν per second. In Fig. 50 is represented a train of waves proceeding from the source S. The observer A, who is upon the bridge, will receive ν disturbances per second. Suppose now that the first ship is moving towards the source S with a velocity v metres per second, whilst the second one is moving away. The observers who are upon these ships would, if they were at rest, also receive ν disturbances per second; but since one observer moves towards the source he receives, in addition to the above number, those disturbances which he meets in the distance AB, which is equal in length to v metres. On the other hand, the second observer fails to receive the disturbances which are in

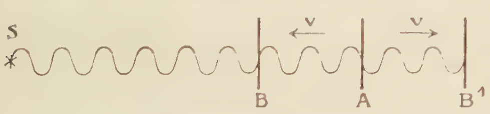


FIG. 50.

the distance AB^1. When all three observers are able to compare observations they will find that each has given a different value to the frequency of the note emitted by S. The observer who moves towards S will hear the highest note, the one who moves away will hear the lowest note, whilst the one who remained upon the bridge will have heard a note the frequency of which lies between those of the other two observers.

If ν is the frequency of the siren, λ the wave-length, then, according to equation (43), we have $\nu = c/\lambda$. This is the number of waves which the observer upon the bridge receives per second, but he receives them some time after they have been emitted from the source; the amount of this delay depends upon his distance from the source, but it will be the same for all waves, *i.e.* it will be equal to the time interval which separates the waves at the source. This case is exactly similar to that of two trains which travel at the same speed but commence the journey at different times. They will pass all points *en route* with a time interval which is equal to the difference between the times at which they set out.

The two observers who are moving with their respective ships will receive $\frac{v}{\lambda}$ waves more [or less]. These waves are those which are in the distance AB [or AB^1]. They will therefore find that

$$\nu' = \nu \pm \frac{v}{\lambda} = \nu\left(1 \pm \frac{v}{\nu\lambda}\right) = \nu\left(1 \pm \frac{v}{c}\right) \quad . \quad . \quad (49)$$

In addition, however, it is possible for the source of sound itself to move. If, for example, a passing cyclist rings his bell to warn us of his approach, or a motor-car with its own characteristic hum passes by, then a perceptible difference in the pitch of the note is to be observed when the cycle or car has just passed by us.

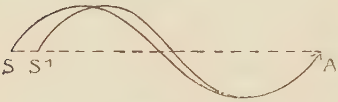


FIG. 51.

Let SA, Fig. 51, be the wave which would be sent out by the whistle of a stationary railway engine. But suppose that, whilst the wave is passing from S to A, the source itself has moved to S', so that the end of the wave begins at S' and not at S. It follows that the wave will have been shortened by an amount SS', which is the distance through which the engine moves in an interval of time equal to the time of one oscillation. If the engine moves in the opposite direction, then the wave-length is increased. If u is the velocity of the engine, *i.e.* the distance through which the source of the sound moves in one second, then the displacement in an interval of time $\frac{1}{\nu}$ seconds is $\frac{u}{\nu}$ cm., and this represents the change in the wave-length, *i.e.*

$$\lambda' = \lambda \mp \frac{u}{\nu}.$$

It therefore follows that the number of waves received by an observer is changed, for from the general equation $\nu = \frac{c}{\lambda}$, we have in the present example

$$\nu' = \frac{c}{\lambda'} = \frac{c}{\lambda \mp \frac{u}{\nu}} = \frac{c}{\lambda\left(1 \mp \frac{u}{\lambda\nu}\right)} = \frac{\nu}{1 \mp \frac{u}{c}} \quad . \quad . \quad (50)$$

If we combine the two cases, so that we have a moving source as well as a moving observer, we obtain the number of vibrations which a moving observer receives from a moving source vibrating with a frequency ν. Thus

$$\nu' = \nu \frac{1 \pm \frac{v}{c}}{1 \mp \frac{u}{c}} \quad . \quad . \quad . \quad . \quad . \quad (51)$$

This equation shows that by a suitable choice of the moving objects a tone of any desired pitch can be produced : it is only necessary to regulate, by the appropriate amount, the number of vibrations which reach our ear per second. It also shows that different observers moving with different velocities will each record a different frequency according to his own particular velocity.

In order to determine the frequency with which a note reaches an observer it is necessary to know the relative motion between the source and the observer : in short, one must know **the number of vibrations which he receives in unit time.**

The Intensity of Sound. Many difficulties confront any attempt to measure the intensity of sound. There is no united opinion as to whether the intensity is proportional to the amplitude or to the energy of the wave, *i.e.* the square of its amplitude. The problem as to how the strength of the sensation varies with the strength of the stimulus also remains unsolved ; and yet, basing our ideas on certain principles, there is no difficulty found in understanding the fact that the sensation will be the more intense, the more intense the effects which act upon our ear from outside.

Quality or Timbre. Finally there is another characteristic which we attribute to sound sensations, and this we call timbre or quality. In general, we can easily distinguish between the note of a violin, flute, trumpet, etc., even when they all three are playing the same note, *i.e.* a note of the same frequency. We say that the three instruments are different in quality.

There still remains, however, another question to be answered. We have to discuss what it is that is present in sound waves which confers upon the different notes their particular characteristics. In an earlier chapter we have seen that, according to **Fourier,** every periodic wave can be represented as a sum of several component waves, the frequencies of which are simple multiples of that of the fundamental. These components are referred to as overtones. If n is the frequency of the fundamental, then $2n$, $3n$ . . . are the frequencies of the overtones. Different instruments will produce notes in which some of the overtones are particularly marked ; one may be rich in the 3rd, 5th, 7th . . . overtones, another in the 2nd, 3rd, 4th . . . ; whilst another may give a note in which the 9th, 10th, 11th overtones are particularly marked. It is due to the varying intensities of these overtones that the waves assume a particular shape, *i.e.* they determine the timbre of the note.

2. THE PROPAGATION OF SOUND

The Velocity of Sound. Generally speaking we may say that a sound reaches us through the air. The air is the most common medium through which sound travels, and one of the necessary conditions for a sound wave to reach us is that the sound shall traverse some medium without interruption, from the source to our ear. A source of sound cannot be heard if it is surrounded on all sides by a vacuum. The intensity of an electric bell is very much diminished when it is placed in an exhausted vessel, although the hammer strikes the dome bell with its usual vigour.

In quite early times it was known that sound required a certain interval of time in order to travel through the air, but a determination of the velocity of sound in air was not made until much later. The first real determination was made by **P. Mersenne** in 1640, although it is true that his results were somewhat high, viz. 450 m. per second. It is easily seen that, in order to measure such a high velocity accurately, the distance over which the sound passes must be large if the interval of time is to be sufficient to permit of the necessary degree of accuracy in its determination. To-day we are able to measure intervals of time as small as 0·001 sec. accurately, but a century ago this was not possible. Experiments were generally carried out in this manner. A gun was fired in the stillness of the night, and the light which accompanied the report indicated the time when the sound began to travel, and the ear was able to decide when the sound arrived. The distance was measured either mechanically or by surveying with theodolites. These early measurements showed that the velocity of sound was not constant. In the first place, it varied with the temperature of the air; in fact the velocity was greater at higher temperatures. The following table contains a few results from recent observations.

TABLE 2

Temperature ° C.	− 140	− 100	− 70	− 40	− 20	− 10	± 0
Velocity m/sec.	226·6	263·0	285·6	306·5	319·3	325·6	331·8
Temperature ° C.	+ 10	30	50	100	200	450	1,000
Velocity m/sec.	337·8	349·6	361	387·2	434	539	717

The greater velocity with increased temperature arises from the fact that the density of the air is then less than at lower temperatures.

Changes in atmospheric pressure do not affect the velocity of sound, for the change in the elasticity of the air is accompanied by a proportionate change in the density.

Newton was the first to derive a theoretical expression for the velocity of sound in air. He obtained the formula

$$c = \sqrt{e/d},$$

which we have already mentioned [*cf.* equation (46)]. The elasticity was calculated in the following manner. The term elasticity signifies for a gas the ratio of a change of pressure or compressing force per unit area to the corresponding change in unit volume of the gas. Now for a gas at constant temperature the characteristic equation is $PV = const.$, *i.e.* the product of the pressure and volume is constant. A small increase in the pressure (p) is accompanied by a corresponding diminution in the volume (v), so that

$$(P + p)(V - v) = const.$$

or

$$PV + pV - Pv - pv = const.$$

But we also have

$$PV = const.$$

Subtracting these two equations from one another and observing that the term pv is the product of two quantities which are small compared with the other variables, so that we are justified in disregarding them, we have

$$pV = Pv$$

or

$$P = \frac{p}{\left(\frac{v}{V}\right)} \quad . \quad . \quad . \quad . \quad . \quad (52)$$

Now the term $\left(\frac{p}{v/V}\right)$ represents the modulus of elasticity [*cf.* equation (36)], for if $p = \frac{F}{a}$ is the force per unit area, and $\frac{v}{V}$ is the change of volume per unit volume, just as $\frac{l}{L}$ was the change in length per unit length. Newton accordingly obtained for the velocity of sound the expression $c = \sqrt{\frac{P}{d}}$. If, however, the numerical values of P and d, each expressed in absolute units, are inserted in this equation, then it is found that $c = 280$ m/sec., a value which is considerably too small. Newton did not succeed in discovering the error in his assumptions. More than a hundred years later, in 1816, **Laplace** solved the difficulty. The formula $PV =$ *const.* which had been used by Newton is only valid so long as the temperature is constant ; Laplace showed that the sound waves follow one another so rapidly, that the variations in temperature which accompany the compression and subsequent refraction of the air cannot be neglected. The effect of this is to require $P\gamma$ instead of P in the above equation, where γ is a definite factor, which, according to thermodynamics, is equal to the ratio of the specific heats of the gas at constant pressure and at constant volume. For air the value of k is 1·4. When Laplace inserted this value in Newton's equation the agreement between theory and experiment was complete.

The Effect of Wind upon the Velocity of Sound. The effect of wind upon the velocity of sound furnishes us with a problem, the discussion of which is very instructive, especially when we come to discuss the manner in which light travels. Whenever a wind is blowing we have a large quantity of air moving at one and the same time. Now since it is possible for motions to take place independently of one another, it follows that sound will travel through a medium in exactly the same way whether the medium is at rest or not. Hence, an observer, in a system where *everything* is moving with the wind, would find experimentally the same value for the velocity of sound in this system, as he would do if it were at rest. An express corridor train provides us with a convenient example of such a system providing that all the doors and windows have been closed. If a person wished to determine the velocity of sound in this system a pistol could be fired in the first carriage whilst in the last he measured the interval of time elapsing between the flash and the arrival of the report. He would obtain a value equal to that which he would find in the open air. In fact, are we not always really in a similar system ? For is not the whole atmosphere moving along with the earth with a velocity of 30 km./sec. ? It is, but the velocity of sound in it tells us nothing at all about this enormous velocity (Newton's Principle of Relativity).

But conditions are changed when we determine the velocity of sound in moving air. Let the starting point of the sound wave be a pistol held in someone's hand whilst the sound is received by another person ; neither of these people moves with the wind, but the sound waves themselves are carried along by the wind. If the direction of the wind coincides with that in which the sound is travelling, then the velocity of the sound is changed, so that if c is the velocity of sound in free air and v is the velocity of the wind, the new velocity of sound under the altered conditions is $c_1 = c \pm v$; the correct sign to be used ($+$ or $-$) will depend upon whether the wind is directed towards, or from, the observer of the sound waves.

The exact velocity of sound, *i.e.* the velocity in still air, can therefore be found as follows: two shots are fired at the same instant from each of two stations and the velocity of the sound proceeding in the contrary directions computed. The mean value of these two velocities is one which is independent of the wind.

Longitudinal Waves. If we now make inquiries concerning the type of wave which is propagated through the atmosphere we shall find that the waves are longitudinal. At the source of the sound waves, *e.g.* the siren, puffs of air are sent out in rapid and regular succession. The result of this is to cause the air which is in the immediate vicinity of the hole to be compressed so that the pressure is greater here than elsewhere. The molecules of air in this compressed air endeavour to move outwards, thereby producing another compression at a little distance from the source. This effect is transmitted to other particles so that the disturbance travels forward. It is also true, however, that a compressed mass of air will exert a force on the molecules of air which lie behind it, and therefore nearer to the source, so that one might well ask the reason why the disturbance is only propagated in one direction. It is necessary to remember that the air which is immediately behind a compression is the same air which recently caused the compression to occur, so that the reaction of the foremost moving air is to cause the air immediately behind it to lose its velocity; an exactly similar occurrence takes place when one moving elastic sphere collides with another one of equal mass, but which is at rest; the reaction of the sphere which receives the impact causes the colliding sphere to come to rest.

There are some cases, however, where the layers of air behind the foremost wave are not brought to rest; these cases occur wherever there is a discontinuity in the medium. A very important example of this phenomenon occurs when a sound wave which has been travelling in the air is incident upon a precipice. The layers of air which we are discussing do not come to rest but move in a direction opposite to that of the original wave, in much the same way as a light sphere moves backwards after colliding with a heavier one. A wave is therefore propagated in the reverse direction and the sound due to such waves is called an **echo.**

Now whilst a siren is in use, the layers of air are only compressed, but most musical instruments cause the air to undergo a rapid alternation of compressions and rarefactions. When the note is sounded for some time a continuous sequence of compressions and rarefactions spreads outwards through the medium in the form of spherical waves. Let us assume that the air at the mouth is suddenly drawn inwards. The air in the immediate neighbourhood is immediately rarefied; in consequence of this the air molecules around this rarefied layer move inwards so that the medium tends to become less dense at this more distant position. This, in turn, affects the air molecules which are further removed from the instrument, so that this rarefaction travels through the air in a manner similar to that in which a compression moves.

It is quite easy to perform an experiment which will show that the air in which sound waves are travelling is undergoing continual changes, *i.e.* the compressions are followed by rarefactions. For this experiment some such contrivance as a sensitive valve is required, *i.e.* we require an instrument in which a valve only permits the flow of a liquid, or gas, in one direction. If a flow tends to take place in the reverse direction the valve will close automatically. Let Fig. 52 represent, on a somewhat large scale, a tube AB, in which the valve can only move outwards towards the left. This means that no air can pass from A to B when one attempts to blow air in this direction through the tube, but air can be drawn in along the tube in the direction B to A. A small glass tube, slightly concave upwards, is attached to the tube AB and it contains a small quantity of coloured petroleum P. On the other side of the tube AB there is fixed a funnel to receive the sound. If now sound waves, the intensity of which is not too low, are incident upon the mouth of this funnel, air can only pass through the valve when suction is applied at the open end of the funnel, *i.e.* where the air is rarefied. This state of rarefaction will be made manifest by the fact that the petroleum tends to move towards B. If the whole contrivance is turned through two right angles so that it faces in the opposite direction the drop P will be moved towards the right; these two facts show that both compressed and rarefied layers of air are present in a sound wave and that these alternate very rapidly.

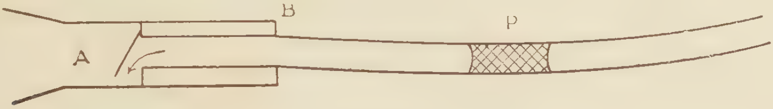


FIG. 52.

The Propagation of Sound in Water. The manner in which sound moves through water is very similar to that in which it moves through air, for water molecules can be displaced relative to one another without forces being brought into play, by the surrounding molecules, which tend to reduce this displacement to zero. It therefore follows that only longitudinal waves, and not transverse waves, can travel in water. These longitudinal waves cause the water to be compressed and then rarefied as in air. Now it is only possible for water to be compressed slightly, *i.e.* forces which are very powerful compared with those in air under similar conditions are necessary in order to compress water even a little; in fact water is only compressed by about 50 parts in a million when the force upon it is 1 kgrm. per cm.^{2} (*i.e.* approximately 10^6 dynes per cm.^{2}). It therefore follows that the modulus of elasticity for water is $c = 10^6/50 \times 10^{-6} = 2 \cdot 10^{10}$. Since the density of water is unity, however, it follows that the velocity of sound in water is c, where

$$c = \sqrt{e/d} = 10^5\sqrt{2} \text{ cm. per sec.} = 1414 \text{ metres per sec.}$$

This is a somewhat surprising result, for although the density of water is nearly 800 times that of air, the velocity of sound in water is only about five times that of sound in air. This is so on account of the enormous forces which are necessary to compress water. The above value has been confirmed by experiment.

The first experiments to determine the velocity of sound in water were made by **Beudant** (1820), who found $c = 1500$ metres per sec. In 1827

Sturm and **Colladon** determined the velocity of sound in the water of Lake Geneva; they gave $c = 1435$ metres per sec. at 8·1° C. A bell was struck below the surface of the water; the same mechanism, which caused this to happen, kindled a heap of powder so that a second person on a distant ship was able to observe the time when the sound waves originated. About ten seconds later he was able to pick up the sound waves by means of a funnel placed beneath the surface of the water.

The Velocity of Sound in Solid Bodies. In solid bodies both longitudinal and transverse waves can be excited. Sound waves travel very easily through such bodies:—the Indians learn of the approach of horse-riders by placing their ears against the earth and listening for the sound of the horses' hoofs.

Let us assume that a tuning fork is allowed to touch one end of a wooden, or metal, rod several metres in length. At the other end of the rod is fixed a resonance box which allows any sound disturbance travelling along the rod to be communicated to the air, Fig. 53. The loud note which is heard under these conditions shows that a sound wave has actually travelled along the rod from the tuning fork. In order to reduce frictional resistances to a minimum it is better to support the rod by means of two thin threads.

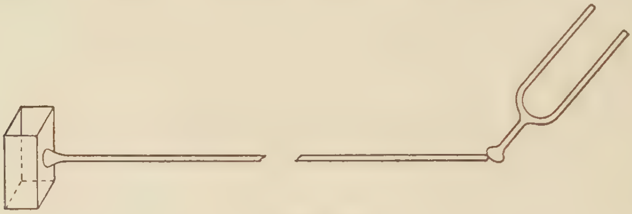


Fig. 53.

The velocity of propagation agrees with that deduced from the formula $c = \sqrt{e/d}$. The following table contains a resumé of the velocity of sound in various gases, liquids, or solids. An inspection of this table shows that, in spite of the great differences in the densities of lead, water and hydrogen, the velocity of sound is approximately the same in all these bodies.

Table 3

The Velocity of Sound (metres per sec.) in various Substances

Air	332	Platinum	2700
Water vapour (96° C.)	410	Pinewood	4800
Hydrogen	1280	Iron	4900
Water	1400	Steel	5000
Stearin	1380	Glass	5200
Lead	1300		

3. Sources of Sound

Preliminary Remarks. As sources of sound waves we must include all contrivances which, irrespectively of the solid, liquid or gaseous state in which they may exist, are capable of producing compressions and rarefactions in the air. But if the sound is to be heard by human ears, then it is essential for the frequency to lie within the range of audibility. If the source of sound is also to be a musical instrument capable of producing different notes according to the will of the player, then the size of the instrument must be such that it is easily controlled and operated by the person who is using it. Fortunately, it so happens that the instruments which give notes within the range of audibility have a size which is convenient for them to be used easily.

Determination of Pitch. The essential difference between the instrument which produces the notes, and the medium which transmits them, lies in the fact that there must be some definite part of the instrument giving one particular note, whereas the medium must be capable of transmitting notes of all frequencies. Now all musical instruments are, of necessity, limited in their length and at their extremities there must be a discontinuity in the medium, and wherever this occurs a retrograde wave is formed. Now every source of sound produces a succession of waves so that the returning waves must interfere with those which are proceeding directly from the source ; *i.e.* stationary waves (*cf.* p. 77) will be produced. The nodes will be situated where the discontinuity occurs at the boundary of a more dense medium, whilst the loops will be formed at the point where the medium becomes less dense. The positions of these nodes, or antinodes, determines the wave-length of the note which a given instrument will produce. The manner in which this wave-length is determined varies in the different musical instruments ; in those instruments which have wires, each end of the wire is attached to a more dense medium whilst, with tuning forks, one end is in contact with a less dense medium (the air). From this it follows that a definite wave-length λ is produced by each particular instrument, and its magnitude bears a simple relationship to that of the instrument. The particular form of this relationship is only to be obtained by a discussion of each type of instrument. In addition to the wave-length, the velocity of propagation is given by the density and elasticity of the material used in making the instrument. From the wave-length and the velocity the frequency of the wave can always be calculated by using the equation $\nu = c/\lambda$. **Hence it follows that in all musical instruments, without exception, only stationary waves can be produced.**

From the equation $\nu = c/\lambda$ we see at once what factors must be changed in order to alter the pitch of a note ; we see that it is necessary either to alter the velocity of propagation or else the wave-length. Now the velocity depends upon the elasticity and density of the vibrating body, whilst the wave-length, as stated above, depends upon the size of the instrument. The means which are at our disposal when we wish to produce a note of definite frequency are therefore as follows : we may change the elasticity, density, or length of the material. There are no other ways in which this can be done, but all these ways can be utilized in practice.

In order to increase the frequency of a stretched wire, which is otherwise very easily displaced, a tension is applied to it ; in this way the frequency of the note emitted can be increased several times. The notes of a piano, violin, or any other form of instrument with wires, are produced by changing the tension in the wires.

In order to change the density of a wire, or rather its mass per unit length, the usual practice is to surround the wire with a spiral of copper so that it shall emit a more sonorous note. Here one may quite well ask why the same effect could not be produced by decreasing the tension in the wire. This could of course be done, but in a musical piece it is essential for the lower notes to harmonize with notes of higher frequency ; besides, it is necessary for all the notes to be of equal intensity. A wire which is only pulled by a small force cannot transmit much energy to the air. Hence, in loading the wire as above, and applying a greater tension, a deeper note can be produced which will sound well with all the other notes.

In actual practice it is found that the third means of changing the frequency is generally employed, *i.e.* the length of the vibrating body is changed. On most occasions this change in length is made by the artist himself; the violinist places his fingers at different points of the string in order to change the pitch of the note. Similarly, with zithers, guitars, and other instruments of this type. In wind instruments their lengths are also altered in order to vary the pitch. This is accomplished in such instruments as the trombone by sliding one tube inside another. In horns, trumpets, etc., the length of a column of vibrating air is changed; in wooden instruments such as the flute, clarionet, oboe, etc., the column of air is effectively shortened, or lengthened, by opening and closing, either with the fingers or by stops which are provided, small holes placed at suitable distances apart. When an organ is being played each note is produced by its own pipe so that an organ is a perfect musical instrument. Each pipe, by making it of the appropriate length, is constructed to give one definite note.

The Excitation of the Air. It now becomes necessary for us to discuss the means whereby these stationary waves transmit their energy to the surrounding air. The frequency of the wave in the air is equal to that of the source, and in order for the wave to be transmitted, the vibrating particle of the source must be in direct communication with the air. The intensity of the note which one hears is not alone determined by the amplitude of the source, but rather depends upon whether or not the vibrating body can cause the air to vibrate powerfully.

In instruments which are blown, it is the air in them which vibrates, and this air is able to transmit its energy to the surrounding air very forcibly and directly, *i.e.* no special contrivances are necessary in order for the sound to be transmitted.

Resonance Boards. On the contrary, however, a vibrating string only comes into direct contact with the air along its breadth so that it only transmits a small portion of its energy to it. In order to increase the amount of energy which is transmitted the string is connected to a resonance board which generally assumes the form of a box made of thin wooden boards. Two bridges are erected at some distance apart on top of the box, the string or wire being stretched across them. When the string is pressed downwards it exerts a greater pressure upon the two bridges, so that these press more strongly upon the resonance board. This is set in vibration with a frequency equal to that of the vibrating string, and on account of its greater area the board is able to act more vigorously upon the surrounding air than the wire alone could do.

At this stage it may be asked how the above process is to be reconciled with the laws of energy, when, by the mere addition of a board which does not do any work, the intensity of the sound can be so much increased. The answer is to be found in the fact that the increase in the intensity of the note is produced by the more rapid and vigorous transmission of the energy from the wire to the surrounding air, the consequence of which is that the vibrating body is brought more quickly to rest than when it is not connected to the resonance board. In 1880 **R. König** made an experiment with tuning forks which were connected to a resonator. When the resonator had a natural frequency different from that of the tuning fork, it did not pick up the energy, and the fork continued to vibrate for 80 seconds. When, however, the resonator was tuned to the same note as the fork, it resounded very loudly, but for 8 seconds only. The string of a violin must therefore be depressed a little more, or else the player must exert himself for a longer time when it is connected with a resonance board than when it is not so connected.

Let us now discuss the different sources of sound individually as far as is deemed necessary by our present requirements.

Vibrating Strings. At either end the strings must lie upon a fixed support, which forms a denser boundary, so that nodes must be formed at the two supports. The length of the string l can therefore be that of a half wave-length, or several half wave-lengths, if intermediate nodes are also produced. The tones which can be produced by a string of length l are therefore conditioned by the following relations :

$$\begin{array}{llll}
l = \frac{\lambda_1}{2} & \therefore \lambda_1 = 2l & \text{and} & \nu_1 = c/\lambda_1 \\
l = 2.\frac{\lambda_2}{2} & \lambda_2 = 2\frac{l}{2} & ,, & \nu_2 = 2\nu_1 \\
l = 3.\frac{\lambda_3}{2} & \lambda_3 = 2\frac{l}{3} & ,, & \nu_3 = 3\nu_1 \\
l = 4.\frac{\lambda_4}{4} & \lambda_4 = 2\frac{l}{4} & ,, & \nu_4 = 4\nu_1 \\
\cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot & \cdot & \cdot\ \cdot\ \cdot \\
\cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot & \cdot & \cdot\ \cdot\ \cdot
\end{array}$$

This whole series of tones are called natural tones.

In Fig. 54 three of these waves are shown. The vibrations of a string are always transverse in their nature and they are produced either by striking the strings (piano), by plucking them (zither, harp) or by means of bowing (violin, cello). This latter is a very remarkable process. A bundle of stretched hairs from the tail of a white horse are rosined, to enable them to grip the strings better, the bundle then being drawn across the string. The string is carried somewhat with the bow, and then having set itself free by suddenly springing away from the hairs of the bow, swings to the other side, and having reached the extremity of its swing on this side, is again gripped by the bow and carried with it. The regularity and purity of the tone depend upon the correct passage of the bow across the string, and this is a matter more of practice than of knowing how.

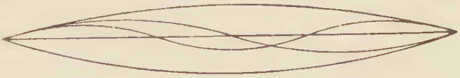


Fig. 54.

Wind Instruments. In these instruments, amongst which the organ is included, a column of air is enclosed in a tube. All such instruments are blown with an interrupted stream of air. The stream is broken in the case of horns, trumpets, etc., by the vibratory motion of the player's lips. With clarionets and oboes the wind from the player causes a thin wooden leaf or reed to vibrate. The motions of the air column act upon this reed so that its own vibrations are in time with those of the vibrating air. With organ pipes the air stream is first blown in front of a sharp edge called the lip of the pipe ; similarly with a flute which is blown towards the rim of the mouthpiece. The action of these lips has not yet been explained clearly. At that end from which the pipes are blown they are all in direct contact with the air. The air which is not enclosed is more rarefied than that which is inside the pipe ; at this end, therefore, all wind instruments have a loop, but the other end can be either open or closed. In most types of wind instrument the pipes are open at the end which is not being blown ; it is only with organs that ‘ stopped ’ pipes are to be found. All the other types must therefore have loops at either end of the pipe and at least one node somewhere along the pipe. If there is one node then the length of the pipe is one half wave-length, Fig. 55 (*a*) ; if there are two nodes, as in Fig. 55 (*b*), $l = 2\,\lambda_2/2$, and when there are three nodes, $l = 3\,\lambda_3/2$, *cf.* Fig. 55 (*c*). Hence it follows that a whole series of tones is again produced as when a wire vibrates ; but with organ pipes it is mainly the fundamental which is used, whereas with trumpets and similar instruments a whole series of overtones, even as far as the eighth, may be utilized. Here it must be mentioned that in Figs. 55 and 56 the waves have been drawn as if they were transverse waves. This has only been done for the sake of clearness, for, in wind instruments, the air, which is the usual medium, can only transmit longitudinal waves.

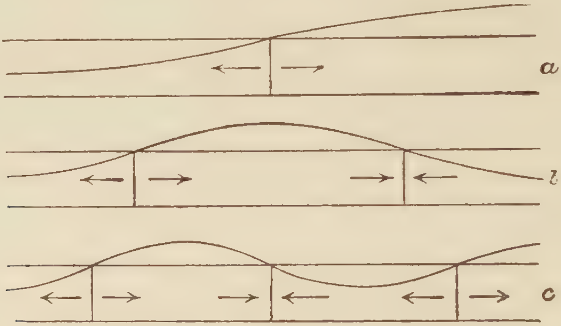


FIG. 55.

With stopped organ pipes a node must be formed at the end which is closed, so that the length of the pipe need only be one quarter of the wave-length it is to produce, *i.e.* the fundamental note has a wave-length $\lambda_1 = 4l$, which is twice that of the corresponding note when a pipe of the same length is open. The organ-builder therefore uses stopped pipes sometimes in order to produce tones which are an octave lower than those from an open pipe of similar dimensions, thereby economizing both material and space. The only overtones which such a pipe can produce are those which have a loop at the open end and a node at the closed end of the pipe, Fig. 56. According to the number of intermediate nodes we therefore have $l = \lambda_1/4$, $l = 3\lambda_2/4$, $l = 5\lambda_3/4$ . . . *i.e.* the frequencies of the upper partials are $\nu_2 = 3\,\nu_1$, $\nu_3 = 5\,\nu_1$, $\nu_4 = 7\nu_1$. . . . The overtones corresponding to the even numbers in this series are lacking. Stopped pipes therefore emit a less rich note.

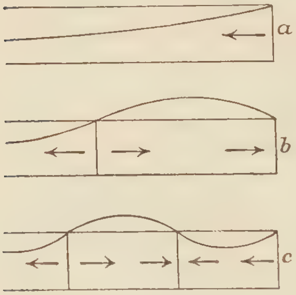


FIG. 56.

Quality or Timbre. The higher tones can be utilized in two ways. They can be added to the fundamental note and give the so-called harmonics, in which case they change the shape of the fundamental wave, as we have already indicated. An untrained ear, it is true, will only hear the fundamental note, but owing to the feebly-sounding harmonics, the note will have a particular quality. Trained ears can, however, easily pick out these overtones. But curiously enough the particular quality of the note is lost to them, and they only hear an additive tone. Or the air column in a trumpet, for example, can so vibrate that although a whole series of nodes is produced, the fundamental is quite absent ; the air does not vibrate with its fundamental frequency. Only the higher notes can be heard. The timbre of different instruments is largely determined by the material of the instrument and its shape. Consequently the player can only affect the quality of a note to a very limited extent.

PART II

THE ATOMIC STRUCTURE OF MATTER

CHAPTER I

THE CLASSIFICATION OF THE ELEMENTS

HITHERTO we have only considered the motions of bodies which could be perceived directly with the eye. But if we see that the laws of nature remain valid whether they are applied to large or small bodies, then we may rightly conclude that the smaller particles which constitute matter are the entities in which the characteristics and properties of matter lie, and that it is from them that we must seek to gain further knowledge concerning the secrets of the material world.

Historical. The ancient Greeks endeavoured to discover the smallest particles which constitute matter, and even they were convinced that an understanding of the greater world could only be obtained by a study of these tiny particles.

Their knowledge of facts, however, was not sufficient to permit them to form a reliable conclusion concerning the atomic structure of matter. We have already learned that the opinions of these ancient philosophers differed very much from one another. They rather tended to discuss all possible alternate hypotheses without ever being able to support one or other of these rival viewpoints by any rigid, or even only convincing, proofs.

The conceptions of **Leucippus** and his pupil **Democritus**, who was born about 460 B.C. at Abdera in Thrace, are the nearest approach, which the ancient Greeks made, towards our modern theories. According to them, all matter is made up of minute particles which cannot be further subdivided, so that they were called atoms. The various atoms differ among themselves only in their size and shape. Composite matter is formed when the different atoms are added together in various ways. Atoms do not fill the whole space occupied by a body completely, for between successive atoms there are large empty spaces. In gaseous bodies the atoms are more distant from one another than in those bodies which are fluid or solid ; this is also true for the atoms of light matter when compared with those of a metal. They assumed that there was an infinite variety of atoms which only differed in size and shape, and not in the substance of which they were composed. Atoms were smaller objects than could be perceived with our eyes and were physically indivisible, although they were not mere points. The cohesion of solid bodies was attributed to the irregular and hook-shaped forms of these minute particles.

If we leave aside these last few remarks, it is surprising to find how near to the truth these ancients were, especially when we realize the scanty and superficial nature of the knowledge which was at their disposal. In their attempt to deduce the enormous qualitative differences which exist throughout the material world, from the quantitative differences which are present in these smallest particles, they were almost two thousand years in advance of their times. As regards the unit of primordial substance, the views of these early workers resemble those of research workers of the present day, who are even now considering in how far known facts substantiate the conception which those men of the distant past announced, without any proof and with almost childlike certainty, to the people of their day.

Unfortunately, **Aristotle**, the greatest amongst the Greek philosophers, did not agree with their views, although he speaks of them in terms of the highest praise, and quite explicitly states that he prefers the methods of Leucippus and Democritus in preference to those of **Plato**. In his teaching concerning the elements, Aristotle follows that of **Empedocles,** who was born in Sicily about 490 B.C., and assumes that the four elements, fire, air, water, and earth, are the primordial substances from which all bodies are formed. But like his teacher Plato, he adds another element —the Æther—in order to account for the heavenly bodies. He likewise assumes that the whole universe is formed from one primary matter, which he called the **materia prima,** which, as such, is nothing, but which can become ordinary matter when it assumes the correct ‘ form ’. The four elements are formed when this primary matter unites with any two of the following opposites—warm and cold, and damp and dry. Thus, the air is warm and damp, fire is warm and dry, water is damp and cold, whilst the earth is dry and cold. The elements can be converted into one another, when one or other of these states is lost, and replaced by another. Aristotle believed that these views were supported by experimental facts. For when a stone or metal is placed in the flame of a candle, the fire is cooled for the stone takes up heat, a new element which is dry and cold, *i.e.* earth, being formed. In fact soot appeared on the stone, and he believed that the fire had been converted into earth. Likewise, when water is heated, he assumed that the cold is driven out from it and the water therefore evaporates. He imagined that the water had been changed into air. Aristotle, however, was far too deep and critical a thinker for the fact to remain hidden from him, that his whole doctrine rested upon an extraordinarily weak foundation, for almost redundantly he himself repeats that, when discussing the greater problems of nature, one must be satisfied with a little success, and if others should succeed in producing more powerful methods and forceful arguments, then our best thanks must be tendered to them. The great attention which the more philosophical part of Aristotle's works has received, was extended to his scientific writings in the following centuries. It was concluded that it was only necessary to infuse the right ‘ form ’ into matter in order to convert worthless bodies into gold. This fact was the motive which caused many people, for more than a thousand years, to seek the Philosopher's Stone, by the aid of which it was thought that this change could be made to take place. It was only after a series of many disappointments and vain efforts that mankind lost its implicit belief in the possibility of causing such a change to occur and realized that chemical research must devise new methods. These new methods must be such that, as had been done in physics for a long time, they were based upon quantitative observations, for it was only by these methods that new knowledge concerning matter could be obtained.

The Elements. In 1777 **Lavoisier,** with a balance at his disposal, showed that mercuric oxide was formed when atmospheric oxygen combined with mercury, and that the same quantities of mercury and oxygen could be obtained again when the oxide was heated. This experiment formed an important landmark in Natural Science. Instead of the continual creation and decay of matter, which was the doctrine postulated by Aristotle, Lavoisier recognized at once an entirely opposite viewpoint. According to him matter cannot be destroyed during any chemical process; *i.e.* he realized the law of the conservation of mass, which has been confirmed again and again with a remarkable degree of exactitude. This doctrine removed the yoke under which science had hitherto laboured, and our knowledge of matter began to progress with amazing rapidity.

In 1799, **Proust** enunciated the Law of Definite Proportions, and **Dalton** discovered the Law of Multiple Proportions, whilst in 1803, the basis of the modern atomic theory, which he had deduced from a consideration of these laws, was expounded by him. According to this theory the ultimate bricks which constitute matter are the minute particles which, following **Democritus**, he called atoms. The whole variety of matter can be traced back to a relatively small number of simple substances, which he called elements. All other substances are formed from these elements when one or more different atoms unite together, and produce the smallest particle of a composite body which can exist, viz. the molecule. In general, only a few different elements are necessary to give rise to a whole range of substances. The plant world is constituted almost exclusively of the three elements, carbon, oxygen and hydrogen. The same three elements in combination with nitrogen are the main constituents of all animal organisms. To-day we know about ninety different elements (*cf.* the tables on pp. 108–9, which give all the known elements with their symbols, names, atomic weights and atomic numbers). We have reason to believe that, except for quite a few, all the elements are now known. The different elements are, according to Dalton, fundamentally different from one another, each being further characterized by an atomic weight, the exact determination of which has been one of the chief tasks of chemistry. **Berzelius** (1779–1848) undertook this study as soon as the idea was conceived and it formed his life's work. With much toil, and a wonderful degree of accuracy, he ascertained the atomic weights of all the elements which were then known, thereby rendering to this new science of chemistry a service which will for ever remain indelibly inscribed in the annals of science.

Very soon after Dalton had propounded his atomic theory **Prout** (1815) extended the theory so that it resembled that of Democritus more closely. He assumed that all the elements were derived from some one kind of primordial matter, and since hydrogen is the lightest of all the elements, he supposed that hydrogen was this substance. This hypothesis was supported by the fact that the atomic weights of many of the elements were integral multiples of that of hydrogen. Further research, however, showed that there were many elements for which this statement was not valid, and so Prout's hypothesis fell into disrepute.

It is now very instructive for us to follow the views which were held at that time. The conclusion, from the fact that the atomic weights were not whole numbers, that it was impossible for hydrogen to be the primary substance from which all bodies were formed, appears so convincing and yet, as we shall learn later, it was false.

The results of chemical research can be expressed in a few words as follows: The material world is composed of about ninety different kinds

TABLE 4

Atomic Weights

(Reproduced with kind permission from the *Journal of the Chemical Society,* 1929.)

Where the last figure of an atomic weight may be in error by several units it is given as a subscript figure. The significance of the last column will be explained later.

Atomic Number.	Name.	Symbol.	Atomic Weight.	Mass-numbers of Isotopes in Order of Intensity.
1	Hydrogen	H	1·0078	1
2	Helium	He	4·002$_2$	4
3	Lithium	Li	6·94	7, 6
4	Beryllium	Be	9·02	9
5	Boron	B	10·83	11, 10
6	Carbon	C	12·003$_6$	12
7	Nitrogen	N	14·008	14
8	**Oxygen**	**O**	**16·0000**	**16**
9	Fluorine	F	19·00	19
10	Neon	Ne	20·18	20, 22, 21
11	Sodium	Na	23·000	23
12	Magnesium	Mg	24·30	24, 25, 26
13	Aluminium	Al	26·97$_0$	27
14	Silicon	Si	28·0$_8$	28, 29, 30
15	Phosphorus	P	30·98$_2$	31
16	Sulphur	S	32·065	32, 33, 34
17	Chlorine	Cl	35·457	35, 37
18	Argon	A	39·94	40, 36
19	Potassium	K	39·10$_5$	39, 41
20	Calcium	Ca	40·09	40, 44
21	Scandium	Sc	45·1$_5$	45
22	Titanium	Ti	47·90	48
23	Vanadium	V	50·95	51
24	Chromium	Cr	52·04	52
25	Manganese	Mn	54·95	55
26	Iron	Fe	55·84	56, 54
27	Cobalt	Co	58·95	59
28	Nickel	Ni	58·69	58, 60
29	Copper	Cu	63·55	63, 65
30	Zinc	Zn	65·38	64, 66, 68, 67, 65, 70, 69
31	Gallium	Ga	69·72	69, 71
32	Germanium	Ge	72·60	74, 72, 70, 73, 75, 76, 71, 77
33	Arsenic	As	74·93$_4$	75
34	Selenium	Se	79·2	80, 78, 76, 82, 77, 74
35	Bromine	Br	79·91$_5$	79, 81
36	Krypton	Kr	82·9	84, 86, 82, 83, 80, 78
37	Rubidium	Rb	85·4$_3$	85, 87
38	Strontium	Sr	87·6$_3$	88, 86
39	Yttrium	Yt	88·9$_3$	89
40	Zirconium	Zr	91·2	90, 94, 92, (96)
41	Niobium (Columbium)	Nb(Cb)	93·3	
42	Molybdenum	Mo	96·0	
43	Masurium	Ma	—	
44	Ruthenium	Rm	101·6$_5$	
45	Rhodium	Rh	102·9	
46	Palladium	Pd	106·7	
47	Silver	Ag	107·880	107, 109
48	Cadmium	Cd	112·4$_0$	114, 112, 110, 113, 111, 116
49	Indium	In	114·8	115
50	Tin	Sn	118·70	120, 118, 116, 124, 119, 117, 122, 121, 112, 114, 115

Atomic Number.	Name.	Symbol.	Atomic Weight.	Mass-numbers of Isotopes in Order of Intensity.
51	Antimony	Sb	$121{\cdot}76$	121, 123
52	Tellurium	Te	$127{\cdot}5$	128, 130, 126
53	Iodine	I	$126{\cdot}93_2$	127
54	Xenon	Xe	$130{\cdot}2$	129, 132, 131, 134, 136, 128, 130, 126, 124
55	Cæsium	Cs	$132{\cdot}8_1$	133
56	Barium	Ba	$137{\cdot}3_6$	138
57	Lanthanum	La	$138{\cdot}9_0$	139
58	Cerium	Ce	$140{\cdot}2$	140, 142
59	Praseodymium . . .	Pr	$140{\cdot}9$	141
60	Neodymium	Nd	$144{\cdot}2_5$	142, 144, 146 (145)
61	Illinium	Il	—	
62	Samarium	Sm	$150{\cdot}4_3$	
63	Europium	Eu	$152{\cdot}0$	
64	Gadolinium	Gd	$157{\cdot}0$	
65	Terbium	Tb	$159{\cdot}2$	
66	Dysprosium	Dy	$162{\cdot}4_6$	
67	Holmium	Ho	$163{\cdot}5$	
68	Erbium	Er	$167{\cdot}6$	
69	Thulium	Tm	$169{\cdot}4$	
70	Ytterbium	Yb	$173{\cdot}_0$	
71	Lutecium	Lu	$175{\cdot}_0$	
72	Hafnium	Hf	$178{\cdot}6$	
73	Tantalum	Ta	$181{\cdot}3$	
74	Tungsten	W	$184{\cdot}1$	
75	Rhenium	Re	—	
76	Osmium	Os	$191{\cdot}_0$	
77	Iridium	Ir	$193{\cdot}0_4$	
78	Platinum	Pt	$195{\cdot}2$	
79	Gold	Au	$197{\cdot}2_1$	
80	Mercury	Hg	$200{\cdot}6_0$	202, 200, 199, 198, 201, 204, 196
81	Thallium	Tl	$204{\cdot}_3$	
82	Lead	Pb	$207{\cdot}2_2$	208, 206, 207
83	Bismuth	Bi	$209{\cdot}0_0$	
84	Polonium	Po	—	
85	———	—	—	
86	Niton (Emanation) . .	Nt (Em)	222	
87	———	—	—	
88	Radium	Ra	$225{\cdot}9_5$	
89	Actinium	Ac	—	
90	Thorium	Th	$232{\cdot}15$	
91	Protoactinium . . .	Pa	—	
92	Uranium	U	$238{\cdot}1_5$	

of elements, the properties of which, and especially the laws governing the manner in which they combine together, have been thoroughly investigated. In particular we have succeeded in comparing the weights of the individual atoms with one another, as the table of the atomic weights shows.

The Classification of the Elements. The Periodic System. In discussing the question of a primary substance from which all elements are composed, another significant problem also arises : Does any incidental relationship amongst the various elements exist ? The Russian scientist **Mendelejeff**, and the German **Lothar Meyer**, almost simultaneously discovered an important regularity amongst the elements (1864). If the whole series of atoms are arranged in order of increasing atomic weight, it is found that their other properties, and especially those of a chemical nature, follow one another. It is also noticed that after a certain number of elements have been passed over, an element is found which is very similar to the first. The next one is similar to the second, etc. The elements can therefore be placed in a series, in which the atomic weight increases as one passes from one element to the next. When an element is found, which is analogous to the first, the series is broken and this element is placed under the first one ; the next is placed below the second, and so on. In this way the **Periodic System of Elements** is obtained, elements which are similar in their chemical properties being found in a group, *i.e.* they are placed under one another. Without exception, all elements find a place in this classification. But it is necessary to consider other properties besides that of atomic weight, so that it is sometimes essential to leave a gap in order for the groups to contain only elements of one class. It was imagined that these gaps represented elements which had not then been discovered, and when some of these elements were discovered and found to possess those properties which had been indicated for them, the importance of this system was at once apparent. When, later on, **Ramsay** (1895 and 1898) discovered a series of new elements which were called the inert gases because they did not combine chemically with other substances, it was found that they constituted a new group in this system. The atomic weights of these elements fitted the system so well that the system received substantial confirmation, at the same time becoming more complete.

In the table on p. 111 the elements are classified according to Mendelejeff. The inert gases are placed in the last vertical column, and since they do not enter into combination with other elements their valency is zero. In the first group lithium, sodium, potassium, etc., are to be found ; they form compounds with those substances which can unite with one atom of hydrogen, so that they are called the monovalent elements. In the second group all those elements are to be found which are able to replace two atoms of hydrogen ; the valencies of the elements in the next two groups are three and four respectively, after which the valency decreases. The elements of the first group are electropositive, *i.e.* they travel with the current during electrolysis and are deposited at the cathode. At the other end of the table the electronegative elements are found, *i.e.* they behave in an exactly opposite manner to the electropositive elements.

But we must not conceal the fact that when an attempt is made to place all the ninety elements according to their increasing atomic weights into this system, some anomalies are found. Several times one has to depart from this normal order if the elements are to be placed in the group to which they belong according to their properties. Thus, the atomic weight of argon is 39·9, a value which is somewhat higher than that of potassium, 39·1 ; nevertheless it is placed before potassium in the table, a fact which demonstrates that all the chemical properties of an element are more important than its atomic weight alone. We shall see later how these anomalies have been explained. For the present we have to be satisfied with the knowledge that we no longer have a mere *number* of elements, but a *system* of elements. In this system there are a fixed number of substances with definite properties, and not an indefinite number of substances with indefinite properties.

Since the atoms of one element always have the same atomic weight, whether they exist as individual atoms, or in combination with other elements, we are compelled to conclude that atoms are fixed entities

TABLE 5

The Periodic System of Elements

Period	Group I		Group II		Group III		Group IV		Group V		Group VI		Group VII		Group VIII	Group 0
	a	*b*	*a*	*b*	*a*	*b*	*a*	*b*	*a*	*b*	*a*	*b*	*a*	*b*		
I	1 H 1·008															2 He 4·00
II	3 Li 6·94		4 Be 9·02			5 B 10·83		6 C 12·00		7 N 14·008		8 O 16·000		9 F 19·00		10 Ne 20·18
III	11 Na 23·00		12 Mg 24·30			13 Al 26·97		14 Si 28·08		15 P 30·98		16 S 32·07		17 Cl 35·46		18 A 39·94
IV	19 K 39·10		20 Ca 40·09		21 Sc 45·15		22 Ti 47·90		23 V 50·95		24 Cr 52·04		25 Mn 54·95		26 Fe 55·84 27 Co 58·95 28 Ni 58·69	
		29 Cu 63·55		30 Zn 65·38		31 Ga 69·72		32 Ge 72·60		33 As 74·93		34 Se 79·2		35 Br 79·92		36 Kr 82·9
V	37 Rb 85·43		38 Sr 87·6		39 Y 88·9		40 Zr 91·2		41 Nb 93·3		42 Mo 96·0		43 Ma —		44 Ru 101·7 45 Rh 102·9 46 Pd 106·7	
		47 Ag 107·88		48 Cd 112·4		49 In 114·8		50 Sn 118·7		51 Sb 121·8		52 Te 127·5		53 I 126·93		54 X 130·2
VI	55 Cs 132·8		56 Ba 137·4		57–71 **Rare-earths**		72 Hf 178·6		73 Ta 181·3		74 W 184·1		75 Re —		76 Os 191·0 77 Ir 193·0 78 Pt 195·2	
		79 Au 197·2		80 Hg 200·6		81 Tl 204·3		82 Pb 207·2		83 Bi 209·0		84 Po 210		85 —		86 Nt 222
VII	87—		88 Ra 226·0		89 Ac 226		90 Th 232·2		91 Pa —		92 U 238·2					

Rare Earths.

VI 57–71	57 La 138·9	58 Ce 140·2	59 Pr 140·9	60 Nd 144·3	61 Il —	62 Sm 150·4	63 Eu 152·0	64 Gd 157·0	65 Tb 159·2	66 Dy 162·5	67 Ho 163·5	68 Er 167·6	69 Tu 169·4	70 Yb 173·0	71 Lu 175·0

which can pass unchanged from one compound to another. For example, carbon always has an atomic weight, 12·0, whether it exists as part of a gas as in carbon monoxide, CO, and carbon dioxide, CO_2, in liquids such as alcohol, $C_2H_5\,(OH)$, or in solid bodies like chalk, $CaCO_3$, or washing soda, $Na_2\,CO_3$. The main task of chemistry has therefore now been accomplished, for it has determined the atomic weights of all the elements, and discovered the composition of all known substances. The inverse process of synthesis, *i.e.* the formation of compounds from their constituents, has also been practically completed, except in the case of complex organic compounds.

But the ever restless spirit of mankind is no longer satisfied with these results alone. The problems must be discussed more deeply. What are the sizes of atoms, how many atoms are there in a gramme of iron, and what is their absolute weight ? Upon what does the relationship amongst the elements, as indicated in the Periodic System, depend, and what are the forces which cause the various kinds of elements to enter into combination and remain together as compounds ? It is fortunate for us that scientific investigations, especially those of the last few decades, have enabled a satisfactory answer to be given to many of the above questions.

In order to appreciate the answers which have been given, however, it is now necessary to discuss in rather more detail the properties of solids, liquids and gases.

Although the substances we call ours and use in our daily life, and which have become familiar to us because we are so constantly in touch with them, are all solids, whereas we are not so familiar with gases, yet it is from the gaseous state of aggregation that we have obtained the deepest insight into the structure of matter, and from it our most accurate knowledge of atomic laws has been obtained. We shall therefore discuss the properties of gases first of all.

CHAPTER II

THE GASEOUS STATE OF AGGREGATION

1. Fundamental Facts

THE time when we first realized that it was the wind which blew into our face was really a remarkable experience in our lives. Until then we had imagined that we were quite free to move at our leisure through the space around us. Now we know that some unseen force can resist our movements. But such experiences first come to us in our early days, and in so many different ways, that we fail to marvel at them. Nevertheless on such occasions we do learn that the space around us is not completely empty and destitute of all properties. When we observe how a storm can bend the branches of a tree, or even break them, we realize that the air is only another form of matter. Moreover the act of breathing makes us conscious of the existence of the atmosphere around us.

The Fundamental Properties of Gases. The chief characteristics of gases which we have to notice at once are as follows :

1. In comparison with solids and liquids, gases are extraordinarily light. The ancient Greeks had emphasized the fact that **air** is light. By this term, however, they understood something which was very different from our modern ideas. To-day the terms light and heavy have only a relative meaning ; a light body is one whose density is low. By the term light body, however, the ancients implied one which tended to move away from the earth, whereas heavy bodies were those which tended to fall towards the earth. Or, according to the doctrines of Aristotle and his followers, light bodies were those whose natural place was beyond that of solids and liquids. The true fact that all bodies possess heaviness, is, however, very old, even though, under the domination of the Aristotelian school in the Middle Ages, it could not assert itself. About 270 B.C. **Straton,** the second successor of Aristotle as leader of the peripatetic school in Athens, taught that gases do not tend to move upwards because they belong to the higher regions, but because they are urged in that direction by the heavier bodies around them. In so doing Straton repudiated the doctrine of natural places, and also declared weight to be a property inherent in all matter.

The fact that gases possess weight is easily demonstrated. If a large glass vessel is furnished with a stop-cock so that it can be closed, it can be proved, even with a moderately sensitive balance, that the vessel is less heavy when the air is removed by a pump, or even partially removed by applying suction with the mouth. Experiment has shown that a litre of air at 0° C., and under normal atmospheric pressure, has a mass of 1·29 grm. Since a litre of water has a mass of 1000 grm. it follows that air is 775 times less heavy than a corresponding volume of water. At 9° C. the mass of a litre of air is 1·25 grm. when it is 800 times less heavy than water. [This latter statement is one which is convenient to be remembered.]

2. The second characteristic of air is the ease with which it can be set in motion, *i.e.* it is very mobile. When the air is only moving with a very small velocity we do not experience any appreciable resistance; it is only when the velocity is considerable, or delicate instruments are used, that a resistance can be felt or detected. The resistance, however, does not make itself felt in such a manner that it tends to reverse the motion, but only acts as a check upon it, *i.e.* it is not like a stretched elastic cord from which a ball is suspended and which endeavours to draw the ball upwards again until the cord is not extended. In other words, the individual molecules in gases are not constrained towards any fixed point. All the molecules enjoy a certain degree of independence. The whole mass of a gas, however, is not a body of fixed form, but:

3. In a gas forces are present which tend to separate the individual molecules from one another. A mass of gas can only be retained if it is surrounded on all sides by walls which are impermeable to it. The apparatus shown in Fig. 57 consists of two glass vessels separated from one another by means of a stop-cock. If one vessel is exhausted, and the other filled with air or gas, a hissing noise is heard when the tap is opened. The noise is caused by the gas which rushes into the previously exhausted vessel, and this flow of gas does not cease until the whole vessel is filled with gas under the same pressure at every point. Gases, therefore, have no volume of their own, but, on the contrary, they tend to fill uniformly the whole of the space which is available to them.

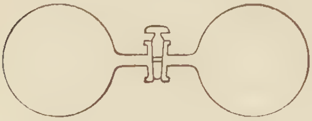


FIG. 57.

The Pressure of Gases. By further investigations on the behaviour of gases an Irishman named **Robert Boyle** (1660) was led to the conclusion that air is elastic, *i.e.* when the pressure on a gas is increased the volume of the gas is diminished until the internal pressure of the gas is equal to that which is applied from the outside. In fact it was he who discovered the simple law which states that **at constant temperature, the product of the volume and pressure of a given mass of gas is constant,** *i.e.* $PV = \text{const.}$ In 1679, **Mariotte** of Paris formed the same conclusion about gases, so that the law is often known as that of Boyle-Mariotte. It is now necessary to consider what the term ‘pressure of a gas’ implies. Let us assume that a mass of gas is enclosed in a cylinder, Fig. 58. The cylinder is closed by a light frictionless piston which, when pressed downwards by a force F, will sink until the volume of the gas is so reduced that the pressure of the gas is equal to that exerted by the piston. The force F distributes itself uniformly over the surface of the piston, so that if this is S cm.^{2}, the force per unit area is F/S, and this force per unit area is called the pressure of the gas. Hence, if P is the pressure, $P = F/S$. A mass of gas enclosed in a vessel exerts the same pressure everywhere, except when the volume of the gas is very large. Under these latter conditions the pressure in the lower regions of the gas is greater than elsewhere by an amount equal to the weight of the superimposed column of gas.

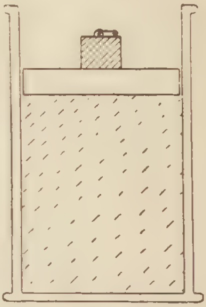


FIG. 58.

Since, according to Boyle's Law, the density of a gas is proportional to its pressure, it follows that the mass of a given volume of gas will increase as the pressure to which it is subjected increases. It therefore follows, that if one moves one metre upwards, the decrease in pressure will be greater the greater the total pressure of the gas, *i.e.* the decrease is greater at low altitudes than at high altitudes, where the gas is no longer compressed to the same extent. The change in pressure which occurs when one ascends can be calculated as follows. We assume that there is a vertical column of air, the cross section being 1 cm.^{2} At any optional height let the pressure be p_1; h metres higher let the pressure be p_2. The decrease in pressure is equal to the weight of the column of air h metres high. If the mass of this air column is m, its weight is mg, so that the decrease in pressure is given by the equation $p_1 - p_2 = mg$. When h is not too large it is always justifiable to assume that the density of the gas in this column of height h is constant everywhere, so that $m = dh$, where d is the density of the air at the point p_1. Now the density of a gas is always proportional to the pressure to which it is subjected. If therefore d_0 is the density when the pressure is atmospheric and equal to p_0, *i.e.* p_0 is the pressure prevailing at the base of the column of air, then the density at any other point where the pressure is p_1 is given by $\frac{d}{p_1} = \frac{d_0}{p_0}$, *i.e.* $d = \frac{p_1 d_0}{p_0}$. Introducing this value into the above equation we have

$$p_1 - p_2 = \frac{p_1 h g d_0}{p_0}$$

or

$$\frac{p_2}{p_1} = 1 - \frac{d_0 g}{p_0} h \quad . \quad . \quad . \quad . \quad (54)$$

This equation implies that, as often as the increase in altitude is h, then the fractional decrease in pressure, expressed in terms of the initial pressure, is always the same. If therefore one rises to various positions, each at a height h above the one immediately below, we have the following equations, if $\frac{d_0 g}{p_0}$ is represented by A.

$$\begin{aligned} p_1/p_0 &= 1 - Ah \\ p_2/p_1 &= 1 - Ah \\ p_3/p_2 &= 1 - Ah \\ &\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ &\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ p_n/p_{n-1} &= 1 - Ah. \end{aligned}$$

Multiplying all these equations together we obtain

$$\frac{p_n}{p_0} = (1 - Ah)^n.$$

By means of this formula we can calculate the change in pressure when the change in altitude is nh. If $H = nh$, then, writing p_H instead of p_n, we have

$$\frac{p_H}{p_0} = \left(1 - \frac{AH}{n}\right)^n.$$

In deriving this formula it has been assumed that the density of the air is constant when we move a distance h upwards. This is only true when h is small. If the increase in height is only 1 metre, then the change in density is scarcely perceptible.

The formula becomes more exact when h is made smaller, and it becomes perfectly true when h is made indefinitely small. This exact formula is easily deduced. If h becomes smaller and smaller (or n larger and larger), then the true value of the pressure at a height H is given as the limiting value when $n \rightarrow \infty$, *i.e.*

$$\frac{p_H}{p_0} = \lim_{n \rightarrow \infty}\left(1 - \frac{AH}{n}\right)^n .$$

The limiting value of the right-hand side of this equation can be found. It is proved in books on infinite series to be equal to e^{-AH}. We therefore have

$$p_H = p_0 e^{-AH}$$

or, if we take natural logarithms,

$$\log_e p_H = \log_e p_0 - AH.$$

In these equations e is the base of natural logarithms, and

$$A = \frac{d_0 g}{p_0} = \frac{0{\cdot}0013}{(76 \times 13{\cdot}6)} = \frac{1}{800{,}000} \text{ (approx.).}$$

According to the gas laws the density of a gas is proportional to its pressure, and the density is also proportional to z, the number of molecules of gas per unit volume, so that we may therefore write

$$\frac{p_H}{p_0} = \frac{d_H}{d_0} = \frac{z_H}{z_0} = e^{-AH} = e^{-d_0 gH/p_0} \quad . \quad . \quad . \quad (55)$$

2. The Kinetic Theory of Gases

Fundamental Facts. When an attempt is made to form from such experimental facts as these, a coherent view of the nature of a gas, many difficulties are encountered. For, on the one hand, we must ascribe to gases, in common with all other matter, attractive forces, which, according to Newton's Law, are inversely proportional to the square of the distance. When the distances are very small these forces become the so-called capillarity forces. On the other hand, one notices in gases an endeavour on the part of the molecules to separate themselves more completely from one another, and this tendency is by no means limited to such small distances as in the case of solid bodies. This difficulty has been overcome by introducing a new idea into Dalton's Atomic and Molecular Theory. It is assumed that the molecules of a gas are in a state of perpetual motion and that, on account of this motion, they collide frequently with one another and the walls of the containing vessel. In the interval of time between two collisions they move freely across a certain distance according to the laws of inertia. The pressure of a gas is a consequence of an extraordinarily rapid succession of molecular impacts upon the walls of the vessel in which the gas is placed. It would have been somewhat illogical to assume the existence of both attractive and repulsive forces in a gas at the same time. The difficulty was removed when the above ideas were introduced. It was only necessary to assume the existence of a force of attraction between the gas molecule, for the pressure which they exerted was due to the fact that the molecules possessed inertia. Although **Gassendi** (1592–1655), **Boyle** (1627–1691), and especially the elder **Daniell Bernoulli** (1700–1782), had expressed these opinions, and in part developed them, they first received general consideration when they were recorded in a small work by **Krönig** (1856). But **Clausius** (1822–1888) must be considered as the real founder of the kinetic theory of gases, whilst **Maxwell** (1831–1879) and **Boltzmann** developed it so that it assumed some measure of finality.

Having thus introduced the idea of the pressure due to a gas, perhaps this is the occasion on which the subject should be examined a little more closely. We can only regard the pressure of a gas upon the walls of a containing vessel as being due to the fact that the wall is subject to an incessant and constant molecular bombardment, at least if such extraneous disturbances as draughts, fluctuations in temperature, etc., are excluded. At all points on the wall, and even at all points and in all directions inside the gas, the same pressure is always exerted. This pressure is explained on the kinetic hypothesis as the result of a large number of molecular impacts. Since these collisions follow one another quite rapidly, but irregularly, the pressure cannot always have exactly the same magnitude. If the pressure could be measured during very small intervals of time, then these variations would be detected. And if smaller and smaller areas of the wall were to be viewed, places would be found at which the impacts were more, and at others less, than the number required to account for the pressure of the gas. Is this an impossible hypothesis? That is, are there any phenomena in the world which contradict our theory? No. Let us imagine a moveable board which is held in position by elastic forces (Fig. 59). When the wall is hit by a ball *A*, it moves back a little, but is constrained to return by the elastic springs. A second ball produces a similar effect. But if the second ball arrives before the board has returned to its zero position, its direction will be reversed before the wall has had time to regain its normal position. And when the balls follow one another more frequently, the board will have less time to return to its original position, so that the to-and-fro motion of the board will become a gentle oscillatory motion: if the ball become smaller, and the collisions more rapid, it will only be with difficulty that the vibrations can be detected. But the elastic springs are now in a continual state of compression. When, therefore, some effect is due to a large number of small individual processes we perceive it as a phenomenon of quite a different character. What is the reason for maintaining this hypothesis when it appears to be contradicted by facts? It is only because of the limitations of our organs of sense, the tardiness of our eyes, and the insensitivity of our touch. If some of these phenomena could be investigated by some being who was so quick in his powers of observation that he could follow the impacts of the individual balls, then he would find that the pressure was not constant, either with respect to place or time. But since, in the end, all our senses, even when supplemented with artificial aids, are no longer able to perceive these fluctuations, it follows that we are not justified in concluding that, because no vibrations can be detected, they do not exist. The conception of a gas pressure being occasioned by a rapid succession of molecular impacts is *not contradicted* by experience.

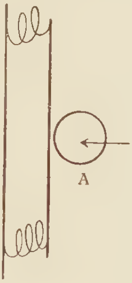


Fig. 59.

But we must not overlook the fact that, by introducing this hypothesis, Boyle's law, which states that $PV =$ constant, when the temperature is fixed, is fundamentally changed: in fact the change is twofold. It is changed with respect to the **interpretation** we give to it, and with regard to its **truth**. First let us consider the interpretation of the law. When the tyre of a bicycle wheel is completely inflated, it becomes quite hard. If an attempt is made to press it inwards, a resistance, such as is usually associated with rigid bodies only, is felt. This resistance is due to the fact that the rubber receives on its inner surface two or three times as many impacts per second, as it does on the outer surface.

Some time ago we referred to the elasticity of air. The idea arose out of the following fact. When a piston, which can move easily inside an air-filled cylinder (*cf.* Fig. 58, p. 114), is made to sink by placing a heavy mass upon it, it experiences a greater resistance as it falls. The resistance is proportional to the density of the gas, and conditions are just as if, instead of an enclosed mass of gas, a spring, in compression, was operating against the piston. But **in the gas there is no vestige of any elastic force,** but only molecules in motion. By diminishing the volume, in which the molecules are free to move, these molecules are pressed more closely together, so that the number of impacts which they make per unit time is increased. If the piston is without weight, and not loaded, equilibrium would be maintained when the number of impacts on its lower surface (due to the molecules inside the cylinder) was equal to the number on its upper surface (due to the external air molecules). If the piston is then raised, the density of the gas inside the cylinder is less than that of the external air—the number of impacts on the lower surface is consequently less. When the piston was released it would behave just like a spring, and execute a to-and-fro motion about its zero position. But the difference between this type of elasticity, and that which exists in a spring, is easily shown by completely removing the external air : the piston would then move outwards until the pressure of the gas inside the cylinder had become infinitely small.

The condensations and rarefactions, which occur when a sound wave is propagated through the air, are produced in somewhat the same way. The effect of the source is to cause a greater or smaller number of air molecules to be accelerated with respect to the layers of air which are more remote. In virtue of this, condensations and rarefactions occur alternately. Let us consider the skin of a drum which is being played. When the skin moves inwards, *i.e.* away from the molecules which are about to impinge upon it from the outside, less molecules will be reflected from it, than under normal circumstances. A somewhat rarefied layer of air then exists above the skin. But when the skin moves outwards more molecules will be reflected from it, than when it is at rest. The entire propagation of sound occurs in the same manner, and in such a way that the particles successively collide first in denser and then in less dense masses on the more remote layers of the gas.

The kinetic hypothesis enables us to give an entirely different meaning to Boyle's law. When it was originally expounded it was merely a simple summary of observed facts concerning pressure and volume and, as such, over the range then investigated, appeared to be true without exception. Nowadays it has become what is generally known as a *statistical* law of Nature. By means of a large number of events, each independent of one another, and quite irregularly at any one point in space, an essentially new phenomenon appears. The law which governs this new phenomenon states that, on account of the almost infinitely large number of ‘ elementary ’ events, the number of impacts can, at no time or place, differ appreciably from the average value. But one realizes quite clearly that this certainty is very different than when we deduce a law from experimental observations only. In such a case, whether or not an exception will be found, depends upon the fineness of our senses, or the delicacy of our instruments. Small deviations will occur quite frequently, but the greater a deviation is the less frequent will be its occurrence. And even if an appreciable discrepancy should be noticed on some occasion, when the usual instruments are being employed, that would not invalidate the meaning of the law. In place of certainty, a statistical law is really an expression of **probability**. 'It is improbable' that, for any one complete second, the number of impacts made at any one point on the wall should differ considerably from the normal number. The number of impacts made in different seconds will never be exactly the same, but large deviations which are so large in magnitude, and which last for such a time, that they could be perceived by us, occur so rarely, that actually nobody has ever observed them. In the course of our reading we shall often meet such statistical laws, especially when we deal with atomic phenomena. Later on we shall learn about some processes, in which the means of observation have been so much refined, that, by one method, we are able to observe the individual occurrences, whilst by another the total effect is examined.

Theoretical Deduction of Boyle's Law. It is now necessary for us to deduce Boyle's Law as a consequence of the fact that the pressure of a gas must be attributed to molecular impacts. Let us assume that we have a cubical vessel containing a mass of gas. The gas shall be such that all the molecules have a mass m. Let N be the total number of the molecules. The particles will be moving in all possible directions with differing velocities. There will, however, be some mean velocity, of such a magnitude, that if all the molecules had this velocity, the pressure upon the walls would be exactly the same as it is now when the velocities are all different. This mean velocity will be denoted by v and, for the present, will be used in all our calculations.

Since we observe that the motion of the molecules of a gas does not cease on its own account (for otherwise the pressure would gradually decrease and eventually become zero) we have to assume that there is no loss of kinetic energy when a collision occurs; this can only be so if the collisions are perfectly elastic. We shall therefore assume that the collisions are elastic. Hence, in the inner parts of a gas, the collisions are like those of colliding spheres, whereas at the walls the laws governing the collision are those which are valid for the impact of a sphere on an infinitely large wall.

The particles will collide in all directions with the wall, will be reflected and move towards an opposite wall; or it may be that before they collide with the second wall one or more impacts will have been made with other molecules. How shall we be able to introduce law and order into this realm of chaos? It is done by means of a few simplifying assumptions which we are able to make in the case of a gas. The strict mathematical treatment of this subject by **Boltzmann** and **Maxwell** has shown that these assumptions are legitimate.

1. The first assumption has already been mentioned; the molecules are all supposed to have the same mean velocity.

2. We assume further that the different impacts are not made at all angles of incidence upon the wall, but that the molecules are divided into three equal groups each containing $N/3$ molecules, and that the molecules in each group move to and fro parallel to one of the edges of the cube.

3. We also assume that the molecules do not collide with one another in their to-and-fro motion from one wall to the next. This is justifiable because when equally big elastic spheres collide they merely exchange their velocities [*cf.* p. 70]. When therefore we do not have to consider

which sphere moves in a certain direction, but only how many, we can neglect the impact and assume that the two spheres move in their own original directions as if no impact had occurred.

If a denotes the length of the edge of the cube a molecule will traverse a distance $2a$ during the time which elapses between successive impacts on the same wall. Since the velocity of the molecule is v these collisions will occur $v/2a$ times per second upon the same wall. At each collision the momentum transferred to the wall will be $2mv$ since the molecules leave the wall after the collision with a velocity v. The total momentum imparted by each molecule to the wall per second is therefore

$$2mv.\frac{v}{2a}=\frac{mv^2}{a}.$$

Since $N/3$ such molecules collide on the wall, the total transfer of momentum per second is

$$\Sigma ft = Nmv^2/3a.$$

In our problem the impacts occur so regularly that the force is always the same in magnitude. Let this average force be F. Then the total impulse is Ft, and since $t = 1$ (for we have considered the number of impacts per second) we have

$$F = Nmv^2/3a.$$

This force is distributed over an area a, so that the force per unit area, *i.e.* the pressure on the wall, is

$$P = \frac{F}{a^2} = Nmv^2/3a^3,$$

and if we put $V = a^3$,

$$PV = \tfrac{1}{3}Nmv^2 \quad . \quad . \quad . \quad . \quad . \quad (56)$$

It is true that this equation has been derived for a gas contained in a cubical vessel, but the law is valid for a vessel of any shape. The proof is easily extended as follows. Let us assume that we have two equal cubes, each filled with the same gas and each containing the same number of molecules. If the two cubes are placed one on top of each other (Fig. 60) the pressure on the walls cannot thereby be changed. Each side of the wall AB will experience the same pressure since each is receiving the same number of impacts per second. It can now be imagined that the wall has been removed. Under such conditions the molecules, which would otherwise have collided with the wall, will collide with other equal molecules moving with the same velocity. They will therefore return with the same velocity as if the wall were present. The whole vessel will therefore be filled with $2N$ molecules, but since the volume has also been doubled, the pressure will remain as before, even if N is now used to denote all the molecules which are present, and V is the total volume.

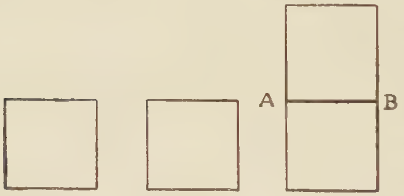


FIG. 60.

Now since any number of cubes can be added together in this manner as often as is desired, and allowed to assume any direction whatsoever, and since large volumes can be formed from these small cubes, it follows that the formula is always true. Now Nm is the total mass of gas available in the given volume and it is not altered when the pressure is changed. The velocity v only changes when the temperature is varied, so that the right-hand side of equation (56) is invariable when the temperature is constant. We therefore have, under such conditions

$$PV = \text{const.} \qquad (57)$$

This equation is an expression of Boyle's Law.

Equation (56) is the fundamental equation used in the kinetic theory of gases. From it we are able to make other deductions concerning the motion of the molecule of a gas; conclusions which we could not possibly derive from experiment alone. Let us write the equation in the form $P = Nmv^2/3V$, where Nm/V is the mass per unit volume. But since we are able to measure the mass of the gas, and also its volume, the value Nm/V is nothing more than the density of the gas. We know, for example, that a litre of air at normal temperature and pressure (0° C., 76 cm. of mercury pressure) has a mass of 1·29 grm., so that its density is 0·00129 grm. per cm.^{3} The equation (56) can therefore be written in the form

$$P = \frac{v^2 d}{3} \qquad (58)$$

In this equation all the variables, except v, which represents the mean velocity of the molecules, are known. This can therefore be calculated from equation (58). We have

$$v = \sqrt{\frac{3P}{d}}.$$

Since one atmosphere $= P = 1{\cdot}013{\cdot}10^6$ dynes per cm.^{2}, we have for a molecule of air at normal temperature and pressure (*N.T.P.*)

$$v = \sqrt{3 \times 1{\cdot}013 \times 10^6/0{\cdot}00129} \text{ cm./sec.} = 447 \text{ m./sec.}$$

For other gases in which the density is d' we have, at *N.T.P.*,

$$v' = \sqrt{3P/d'},\ i.e.\ v : v^1 = \sqrt{d' : d} \qquad (59)$$

Since the density of a gas can be determined by observations made with a balance, the mean molecular velocity can be computed. The following table contains values of this velocity in the case of some common gases.

TABLE 6

Molecular Velocities of Some Gases.

Gas.	Formula.	Molecular Weight.	Mean Square Velocity (Metres per Second).
Hydrogen	H_2	2	1692
Helium	He	4	1204
Ammonia	NH_3	17	582·7
Neon	Ne	20·2	535
Carbon monoxide	CO	28	454·5
Nitrogen	N_2	28	454·3
Air	—	—	447
Oxygen	O_2	32	425
Hydrogen chloride	HCl	36·5	398
Argon	A	39·9	381
Carbon dioxide	CO_2	44	362·5
Chlorine	Cl_2	71	285·6
Xenon	X	130·2	211
Bromine	Br_2	160	191
Mercury Vapour	Hg	200·6	170

This table shows that the molecule of hydrogen moves with a velocity which would cause it to travel more than one and a half kilometres per second, that of helium more than one kilometre, whilst for the molecule of air the distance would be half a kilometre. We must now attach some definite meaning to this velocity. It does not imply that every molecule moves this distance from its initial position in each second ; for the molecules collide with other molecules before they have moved very far, and are forced to return, but their velocity is such that if they did not collide with other molecules then they would move uniformly through this distance in one second. If the distance which they do traverse in the interval of time between two collisions is divided by this time, then the average value obtained is equal to that given in the above table.

The significance of the table, however, must not be overestimated. The numbers are neither a confirmation nor an achievement of the kinetic theory of gases—at least not at present. The meaning of the numbers, however, is somewhat as follows : if one wishes to regard the actual pressure of a gas as a result of molecular collisions, then the molecules must possess these velocities, for otherwise their momentum would not be sufficient to account for the observed pressure. But the importance of these members will become very great if one should even once be able to measure the velocities of these molecules by means of experiments not based upon the kinetic theory. Under such conditions the kinetic theory of gases would cease to be a hypothesis, for henceforth it would have to be counted amongst facts which can be observed. We shall soon learn how such experiments have been made and then the numbers recorded in the above table will permit us to obtain a deep insight into the inner mechanism of a gas ; the numbers will then be the actual velocities of the gaseous molecules. They reveal a fundamental property of matter to which our attention was first directed by the kinetic hypothesis.

It has frequently happened in the history of science that a hypothesis has permitted deductions to be made which, on the one hand, were until then quite unknown, but which, on the other, after they had been proposed as questions, could be subjected to direct experimental verification. When this has happened, it has been regarded as a very strong proof in favour of the validity of the hypothesis—it is said that the hypothesis has a heuristic value. We shall now proceed at once to become acquainted with several such deductions.

First of all let us contemplate the exact meaning of the term ‘ mean velocity’. The mean velocity has been calculated from the formula $PV = \frac{Nmv^2}{3}$. The term v has therefore to be chosen so that if v_1, v_2, v_3, . . . , the actual velocities of the different molecules, were replaced by it, the pressure upon the wall will remain the same, *i.e.* v must be chosen so that

$$Nv^2 = v_1^2 + v_2^2 + v_3^2 + \ldots + v_N^2.$$

Hence

$$v^2 = (v_1^2 + v_2^2 + v_3^2 + \ldots + v_N^2)/N.$$

Thus v^2 is the mean square velocity, the square root of which quantity gives the value of v. This value v is not the same as the mean value of the velocities v_1, v_2, v_3, . . . v_N, which is

$$(v_1 + v_2 + v_3 + \ldots v_N)/N.$$

A numerical example will suffice to convince us that this is true. The mean value of 7 and 17 is 12, but the square root of the mean value of the squares, *i.e.* $\sqrt{\frac{7^2 + 17^2}{2}} = \sqrt{169} = 13$.

Maxwell has calculated, in addition to the root mean square value which our equation gives, also the mean linear velocity which the above example shows us is somewhat less than that of the root mean square value. Maxwell proved that it was about 8 per cent. smaller and the ratio of the mean linear, to the root mean square velocity, is 0·9213. This difference must always be considered in any problem which may arise, and according as to whether one is dealing with the mean **energy** or the mean **velocity**, the mean square or mean linear velocity must be used.

Diffusion of Gases. Let us assume that there are two vessels, *A* and *B*, Fig. 61, each filled with a different gas. For example, let *A* contain hydrogen and *B* oxygen, both gases being at atmospheric pressure. The two gases are separated by a porous wall *C*, which is made of an unglazed tile, piece of plaster of Paris, or some similar material. After a little while, temperature and external pressure changes being avoided, some hydrogen will be found in *B*, and some oxygen in *A*. More hydrogen will have passed through the porous wall into the other vessel than oxygen has passed in the reverse direction. The pressure in *B* will therefore have risen whilst that in *A* will have fallen. This phenomenon is referred to as that of the diffusion of gases. If the quantity of gas which can so diffuse in a given time under given conditions is compared with that of another gas under the same conditions, then one is led to the idea of a velocity of diffusion. Experiment shows that the velocity of diffusion is inversely proportional to the square root of the density of the gas. If it were supposed that the gas molecules were for ever at rest, then the phenomenon of diffusion would present many difficulties. But it is now very easy to explain it as a necessary consequence arising from the motion of the gas molecules, and it is still more satisfactory if the different diffusion velocities can be explained. Now, according to the kinetic theory, the velocities of the molecules are inversely proportional to the square root of the density of the gas. It is therefore fairly obvious that the different velocities of diffusion arise because the velocities of the molecules themselves are different.

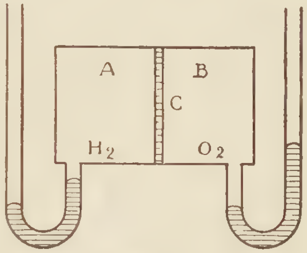


Fig. 61.

Van der Waals' Equation. In 1842 the French physicist **Regnault** made some careful investigations concerning Boyle's Law, and found that it did not represent the facts exactly. The product PV, which at constant temperature should always be invariable, showed small deviations. When he determined the volumes $V_0, V_1, V_2, \ldots$ at pressures $P_0, P_1, P_2 \ldots$ and then calculated the products P_0V_0, $P_1V_1 \ldots$ the ratios $\frac{P_0V_0}{P_1V_1}$, . . . should all have been equal to unity, if the product PV was invariable. With different gases, however, he found small deviations, and from the very careful measurements which he made, he was able to convince himself that these deviations did not arise from experimental errors, but that they represented departures from Boyle's Law. The deviations were very small in the case of hydrogen, but were greater with those gases which can be liquefied at ordinary temperatures. These deviations became more marked at lower temperatures, and increased considerably as the temperature at which liquefaction took place was approached.

At first this discovery appeared to present certain difficulties to the theory, which, until then, had appeared to be in favour of Boyle's Law, expressed in its simplest form. In 1873, however, **Van der Waals** of Amsterdam (1837–1923) proposed a new equation, which was a modified form of the original one due to Boyle, and which would account for the deviations observed by Regnault. He drew attention to a series of omissions which had been made in deducing Boyle's equation, and showed how they could be avoided. In the proof which we have given above, it was assumed that the whole of the space occupied by a gas is available for the motions of the molecules, and that between two impacts on the same wall the molecule would move a distance $2a$. This is not quite correct. The molecules will have a certain diameter of their own—if it is denoted by $2r$, then $(a - 2r)$ is the distance through which a molecule moves in passing from one wall to another. This must be true since the centre of gravity of the molecule cannot approach the walls nearer than a distance r. Also, in a collision between two molecules, the centres must remain a distance r apart. In short the whole space occupied by a gas is not available for the movements of the molecules, but a space which is somewhat less than this by an amount depending upon the volume of the molecules themselves. The difference is actually even greater than the volume of the molecules themselves. Van der Waals calculated that the difference was equal to four times the actual volume of the molecules. He therefore wrote $(V - b)$ in Boyle's equation instead of V, where b is equal to four times the actual volume occupied by the molecules of the gas themselves.

In addition to this, when deducing Boyle's equation, it was assumed that the pressure only depends upon the velocities given by the laws of impact—in other words no account has been taken of any forces which may exist between the molecules. These forces do exist, however, and the more densely the molecules are packed together, the nearer do the molecules approach one another, and the more frequently do the collisions amongst them occur. The force of attraction will therefore increase with the square of the density, so that the pressure, as measured externally, will be too small. Since the molecules themselves have a volume of their own, the pressure will be higher. Van der Waals therefore wrote, instead of Boyle's equation,

$$\left(P + \frac{a}{V^2}\right)(V - b) = \text{const.} \qquad . \quad . \quad . \quad (60)$$

This equation represents the precision measurements of **Regnault** and later observers with much greater accuracy. The values of a and b were calculated from observations on some few values of P and V. The equation was confirmed when it was shown that all other values of P and V were satisfied by these values. The values for a and b as found by Regnault are recorded in the table below. The unit of pressure is that of a column of mercury 1 m. high and the unit of volume that of 1 kgrm. of the gas.

TABLE 7

Gas.	*a*.	*b*.
Air	0·0037	0·0026
Carbon dioxide	0·0115	0·003
Hydrogen	0·0003	0·00069

When, with any particular gas, the deviations from Boyle's Law are not noticeable, the fact is stated more briefly as follows: The gas is a **perfect gas**. This phrase is also applied to gases such as hydrogen, helium, and nitrogen, in which the values a and b are only very small.

This refinement of the original equation due to Boyle is important for our purpose, because it allows us to obtain a further insight into the internal mechanism of a gas. In the quantity b we have one which, although it does not represent the actual volume of the molecules of a gas, is yet related to this volume, so that we can calculate from it the order of magnitude of this quantity. Since the value which is so deduced is confirmed by values deduced in other ways, the significance of b can no longer be doubted.

The attractive forces represented by the term a in Van der Waals' equation must not be regarded as merely forces to be accounted for by Newton's laws of gravitation. In the sections on solid and liquid bodies we shall see that material particles in close proximity exert other forces upon one another, and it is to these forces that we must look when we attempt to account for the cohesion of solids. In fact, they are called the forces of cohesion. These same forces are also to be found amongst the molecules of a liquid, and they are also the same forces which cause deviations from Boyle's law to exist in gases.

The Thomson Effect. The forces which have just been mentioned play a very important rôle in the Thomson effect. **Thomson** (1824–1907) and **Joule** in 1852 discovered that gases are cooled when, after having been compressed, they are allowed to expand under such conditions that no external work is done. This cooling is due to the fact that the molecules must do work against these forces of mutual attraction when they tend to become separated from one another. This loss in kinetic energy appears as a gain in potential energy. Those gases which are generally considered to be perfect gases, like hydrogen and helium, also exhibit this effect when they expand at low temperatures, *i.e.* when they have already been cooled in some other way. This property of gases is used in **Linde's** liquefaction apparatus in order to produce very low temperatures. By the use of this machine hydrogen and helium have been liquefied.

The Kinetic Energy of a Gas. Equation (56) gives us a very remarkable relationship which exists between the energy of the molecules of a gas due to their motion, and their pressure and volume. This relationship must now be discussed briefly. Let us assume that a cylindrical vessel (Fig. 62) is filled with a gas at pressure P. The vessel is closed by means of an easily moving piston. We shall neglect the weight of the piston, and also the friction, which must be present between the piston and the glass walls. If the external pressure is also P, the piston is in equilibrium. Let us suppose that the gas is compressed by introducing more gas through a side tube. If the piston is at first at a somewhat low position in the cylinder, the incoming gas will cause it to rise slowly, a fact which indicates that the incoming gas does work because it overcomes the opposing pressure P. What is the magnitude of this work ? Since the term pressure signifies the force per unit area, the total force which opposes the motion of the cylinder is $F = PS$, where S is the area of the piston. If the piston has risen a distance h when all the gas has been introduced the work done is

$$W = Fh = PSh = PV \quad . \quad . \quad . \quad . \quad (61)$$

Hence PV denotes the work which must be done by the gas against the external pressure, when the increase in volume is V. From whatever source the gas is introduced into the cylinder, whether it has been produced by the chemical action of an acid upon a metal, or by heating a liquid, the work done is always PV, when the piston is moved against the pressure due to the external atmosphere. The product PV can therefore be called the volume energy of the gas.

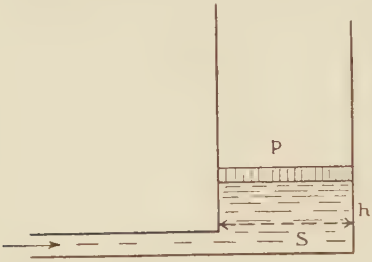


FIG. 62.

The expression on the right-hand side of the equation is somewhat remarkable for it is equal to $Nmv^2/3$. In this, mv^2 is twice the kinetic energy of a molecule, or at least of a molecule which is moving with the mean kinetic energy, and N is the number of molecules ; the total kinetic energy due to the motion of the molecules is therefore $Nmv^2/2$. If this is put equal to E_k, *i.e.* $E_k = Nmv^2/2$, we have

$$PV = Nmv^2/3 = \tfrac{2}{3}E_k \quad . \quad . \quad . \quad . \quad (62)$$

This expression gives us a simple relationship between the volume energy of a gas and the energy of its molecules due to their motion, which is true for all gases. If the volume energy PV is denoted by E_v, we have

$$PV = E_v = \tfrac{2}{3}E_k \quad . \quad . \quad . \quad . \quad (63)$$

The volume energy of a gas is therefore equal to two-thirds of the kinetic energy due to the motion of its molecules.

Avogadro's Hypothesis. After the law of constant proportions in connexion with the formation of chemical compounds had been discovered, the reactions of gases received consideration. It was found that when gases combine with one another completely, the volumes of the reacting gases, under the same conditions of temperature and pressure, are always equal, or else bear a simple ratio towards one another. Thus, one litre of chlorine combines with one litre of hydrogen to form two litres of hydrogen chloride ; whilst one litre of hydrogen and half a litre of oxygen combine to form one litre of water vapour. Now since, according to **Dalton's** atomic theory, only those substances combine which contain an equal number (or a simple multiple) of molecules, one is forced to assume that equal volumes of gases (always measured under the same conditions) contain the same number of molecules. This hypothesis was proposed by **Avogadro** in 1811. He stated that equal volumes of different gases, under the same pressure and at the same temperature, always contain the same number of molecules. It therefore follows that the masses of these equal volumes must be proportional to their molecular weights since, in general, the ratio of the same number of two different things is equal to the ratio of the things themselves. For example, the ratio of the mass of a thousand shillings to that of a thousand sixpences is equal to the ratio which exists between the mass of one shilling and that of one sixpence. The tests which could be made on some known substances often showed agreement, but discrepancies were occasionally observed. These discrepancies were explained when it was discovered that at the temperature of the experiments the molecules of a gas could dissociate. Under these conditions one was no longer dealing with a single gas ; if these instances are omitted, Avogadro's hypothesis has always been confirmed, so that to-day it can be considered as a fundamental law of gases, which has been proved both by direct and by indirect experiment.

CHAPTER III

THE FLUID STATE

1. FUNDAMENTAL PROPERTIES OF LIQUIDS

THE most prominent feature of liquids is the ease with which they can be displaced. If we wish to cause a body to move rapidly through water a certain resistance is felt ; this resistance is not produced by forces tending to maintain the molecules of the water in their relative positions, but is due to the inertia of the masses of water which have to be thrust aside so quickly. But apart from the inertia of the molecules there are forces which must be overcome when a body is moved through water. It is easy to discern the great differences in mobility which exist between ether, water, oil, glycerine, treacle, etc., a list which could easily be extended. Having passed the very sticky or viscous liquids, substances such as pitch and asphalt occur. On a hot summer's day these are almost fluids, whereas in winter they are hard and brittle.

We see, therefore, from the above remarks that it is difficult to draw a sharp line of demarcation between liquids and solids. There are substances which are undoubtedly liquid, whilst there are also other substances to which only some of the properties of a liquid can be ascribed.

The idea of a perfect liquid has therefore been formed. Such an ideal liquid is one in which, apart from its inertia, there are no forces present, which will tend to retard the motion of a body through it. Those forces which do retard the motion of bodies are due to viscosity and are known as viscous forces. **Perfect liquids are therefore those in which the viscosity is zero.**

The Shape of Liquids. One of the immediate consequences of the fact that liquids are mobile is that they do not have any shape of their own, but, since they are subject to gravity, they always fill the lower portions of any vessel into which they may have been placed. When gravity is the only force to which they are subject the upper surface of a liquid is quite smooth ; in small vessels this surface appears to be part of a plane perpendicular to the line of action of gravity. The surface of an ocean is, in the absence of a wind, also perpendicular to this line of action but it is not part of a plane. The surface really forms part of the curved surface of the earth, which is approximately that of a sphere. This lack of perfect symmetry is due to the fact that the water is heaped up at the equator on account of the existence of centrifugal force there, so that an ellipsoid of rotation is a better approximation to the true shape of the earth.

It is easy to prove that a quantity of water can only be in equilibrium when its surface is perpendicular to the line of action of gravity. Let us assume that the shape of a liquid surface has become different from this owing to the existence of some other force (Fig. 63). Suppose that the force which gravity exerts upon a particle at C is represented by the straight line CD. This force can be resolved into two components, one of which is perpendicular to the surface, whilst the other acts along the tangent at C. This latter force must accelerate the particle at C so that it tends to move down an incline. This happens, in a similar way, with all the other particles which are above the level AB. They all flow downwards so that the protuberance vanishes. If all the molecules lie in the plane AB gravity is everywhere perpendicular to the surface whilst the component parallel to the liquid surface is zero.

The Pressure in Liquids. Owing to the mobile nature of a liquid it follows that any pressure which is applied externally to a liquid will be transmitted to all points in that liquid. For a particle which experiences even a slightly greater pressure on one side than on another will, on account of the mobility of the liquid, move in the direction of this excess pressure, thereby tending to increase the pressure at the point to which it moves. This particle will likewise exert a pressure upon its neighbour so that the whole process continues. Finally some particle will collide with a solid wall and exert this same pressure upon it until the reaction which is due to the elastic nature of the wall holds the particle in equilibrium. Or else, if this last particle moves against some opening in the wall it experiences no reaction so that it becomes accelerated and leaves the opening with the corresponding velocity.

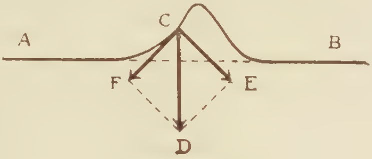


FIG. 63.

The True Volume of Liquids. On account of the ease with which liquids move under gravity, they occupy the space in a vessel without leaving any gaps. Since we are not able to see the individual molecules with our eyes, or even with the most powerful microscope, a liquid appears to us to be absolutely continuous. It is therefore self-evident that the only conclusion we can make from this is somewhat as follows : if discontinuities do actually occur in the space occupied by a liquid, then these must be smaller than anything which can be detected with our eyes even when we are provided with some external aid to vision. In any case it follows from the large mobility of the particles under the influence of attractive forces that the space is very completely utilized and the particles closely packed together, so that the second fundamental property of liquids is that they possess a definite volume.

Bulk Elasticity. If we endeavour to compress a liquid it offers a great resistance to the compressing forces, but after they have been removed it immediately assumes its original volume—in other words, liquids are characterized by a bulk elasticity. On account of this elasticity they are able to transmit longitudinial waves—thus water will transmit sound waves. The velocity of propagation is very high in spite of the great inertia of the water, because the modulus of elasticity is very large. Let us denote again by λ the ratio of the diminution in volume v to the original volume V, *i.e.* $\lambda = v/V$. We have already seen that the equation (35), $e = f/\lambda$, is valid for gases. It is also true for liquids. The contraction which occurs when water is under a pressure of one atmosphere is 50×10^{-6} times its volume. This compressibility varies with the temperature of the liquid. In the case of water this variation is given in the following table.

TABLE 8

Temperature (° C.).	0	10	20	30	50	70	100
Compressibility × 10^6	51	48	47	46	45	46	48

For several other liquids at 0° C. the coefficients per atmosphere are given in the following table, in millionths of the initial volume.

TABLE 9

Æther.	Alcohol.	Acetic Acid.	Oils.	Petroleum.	Mercury.
170	100	40	50–60	70	3·9

Since the compressibility of liquids is so very small, one may again define a perfect liquid as one which is absolutely incompressible. Although there are no such perfect liquids, the deviations found in actual practice are generally negligible.

Cohesive Forces. A whole series of phenomena are known in connexion with liquids, which signify that forces do exist between their molecules. A small quantity of mercury always assumes a spherical shape. Even a drop of water placed upon a piece of paraffin wax, or some other substance which it does not wet, has the shape of a sphere. If a larger quantity of liquid is used (see Fig. 64) the effect of gravity is to cause this shape to be flattened, but the existence of forces tending to maintain the spherical shape is easily observed. If a clean plate of glass Fig. 65, is suspended carefully in a horizontal position from a balance, and allowed to dip into a vessel containing water, a considerable mass has to be placed in the other pan of the balance before the plate will leave the water. What we actually measure in this experiment are the forces which hold the individual molecules together, for since the plate remains wetted by the water, the separation takes place between the water which adheres to the plate and that which is left in the vessel. These adhesive forces are due to capillary attraction.

FIG. 64.

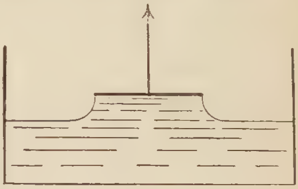


FIG. 65.

These forces are much stronger than the Newtonian forces of attraction so long as the particles are in immediate contact with one another, but as soon as the particles are separated from one another by a distance even less than 0·001 mm., these forces vanish completely. All such phenomena which have hitherto been observed can be explained, without exception, on the hypothesis that there are forces acting between the molecules of a liquid so long as the distance between them is less than a certain amount a, but vanish completely when the distance is greater than this limit. This distance is represented by the radii of the spheres A, B, C, D, shown in Fig. 66—the radius has been enlarged very considerably in order to facilitate the drawing of the figure. At the centre of each sphere there is a molecule. It is only the first molecule *A* which is attracted by all the molecules which lie within its 'sphere of action'. For the molecules *A* and *B*, which lie at a certain distance below the surface of the liquid, so that their sphere of action is entirely within the liquid, this attraction will not be perceived because the effects are the same in all directions, so that the resultant effect is zero. Let us consider, however, a molecule *C* lying on the surface of the liquid so that the upper half of its sphere of action lies outside the liquid where there are no molecules. The molecule *C* will therefore experience a one-sided pull tending to bring it into the liquid. Molecules such as *D*, which are so near to the surface of the liquid that part of their sphere of action will lie outside the liquid, will likewise be drawn towards the interior of the liquid because the forces due to the molecules in the lower half of the sphere will not be completely balanced by the forces due to the molecules in the upper half of the sphere. The excess force in the downward direction will tend to make the molecule move towards the interior of the liquid. In consequence of these forces it is only possible for a molecule to be brought to the surface of a liquid by expending a certain amount of work; conversely, molecules at the surface of the liquid will experience forces tending to diminish the surface area. Hence all liquids which are not affected by other forces assume a spherical shape, because a sphere is the body having a minimum area for a given volume. The spherical shape of the heavenly bodies is to be attributed to these forces, for such bodies were originally liquid, and, having cooled down, this shape has been maintained.

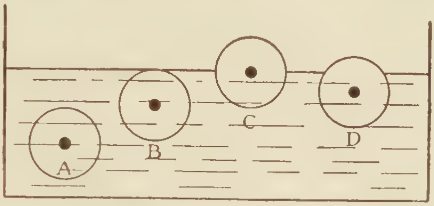


Fig. 66.

Adhesive Forces. Between liquids and solids there are also attractive forces present which only arise when the distance of separation is very small. These are the so-called forces of adhesion. They can even be much greater than the corresponding forces between the molecules of a liquid. When this happens the body is wetted when it is withdrawn from a liquid; but if the molecules of the liquid attract one another more closely than they do those of the body immersed in them, the body is not so wetted. This occurs when a clean glass plate is withdrawn from mercury. Nevertheless measurement shows that mercury adheres more strongly to glass than does water. It is owing to the relationship which exists between cohesion and adhesion that water rises at points where it touches the walls of a vessel, or in a capillary tube immersed in the surface of the liquid, Fig. 67, whereas mercury is depressed, Fig. 68.

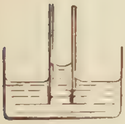


Fig. 67.

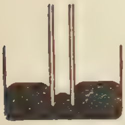


Fig. 68.

On the whole our knowledge concerning the nature of these forces is remarkably deficient and, in particular, this subject has hitherto contributed very little towards our understanding of the molecular structure of the material world.

2. Molecular Movements in Liquids

We have reason to suppose that the molecules of a liquid are in a state of perpetual motion. On account of the more dense packing of the molecules in a liquid, however, the distance traversed between successive impacts will be much shorter than in gases. Consequently the molecules in a liquid will not move very far from their mean position. The fact that such a state of molecular motion does exist will be proved by a whole series of phenomena.

Diffusion. First of all liquids exhibit the phenomenon of diffusion. If a blue solution of copper sulphate in water is placed carefully as a layer at the bottom of a vessel containing pure water, a definite line of demarcation is visible at first. If the vessel is protected from extraneous disturbances, it will be found after a few days that the demarcation line is not so definitely fixed ; in spite of its greater density some of the copper sulphate solution will have travelled upwards and mixed with the water. After a much longer time, perhaps several weeks, it will be found that there is no line of demarcation between the upper and lower portions of the liquid. The heavier sulphate solution will have penetrated everywhere so that the concentration will be the same at all points. These phenomena would be very difficult to explain if one had to assume that the particles of a liquid were completely at rest relative to one another ; but it results as a necessary consequence if one assumes, as in the case of gases, that the molecules are in a similar state of continual motion.

Osmosis. From the fact that the dissolved molecules of copper sulphate have moved upwards through the water, and thereby performed work against gravity, it follows at once that there must be present in a solution a force tending to make the concentration uniform. If the particles can yield to this force, then it is revealed by a motion of the particles, even against gravity. But if the motion of the molecules is checked, then they must exert a pressure upon the body which is holding them back. This pressure is called the **osmotic pressure** of the solution. Phenomena due to osmotic pressure have been known for quite a long time. The first observations were made by the Abbé **Nollet** of Paris in 1848, but the first exact measurements were made by the botanist **Pfeffer** in 1877. He had succeeded in making some so-called semi-permeable membranes. These are thin membranes which allow the water in a solution to pass through them, whereas they are not permeable to the dissolved. Hence the membrane must experience a pressure tending to maintain equilibrium.

Animal membranes are not perfect, when used for this purpose. A membrane, which is perfect for many substances, is obtained by pouring copper sulphate solution into a porous pot, and then dipping the whole in a vessel filled with a solution of potassium ferro-cyanide. At the junction of the two liquids, which occurs inside the walls of the porous pot, a precipitate of copper ferro-cyanide is formed; this allows water to pass through it, but not dissolved salts and other substances. If such a prepared pot is filled with a solution of sugar and placed inside a vessel of pure water, Fig. 69, the water at once begins to pass into the cylinder so that the pressure therein is increased. After a time the pressure attains its maximum value, which then remains constant. The total change in pressure is the osmotic pressure of the solution, and it is measured by the height of the mercury column *EF*, which is in equilibrium with it.

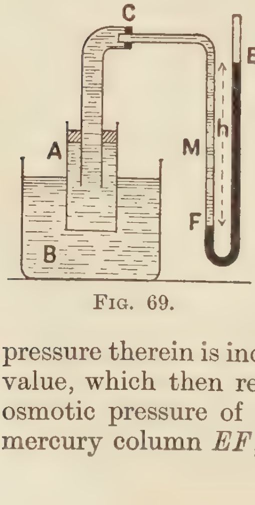


Fig. 69.

Pfeffer found that, at constant temperature, the osmotic pressure of a solution is proportional to its concentration. The most remarkable of his discoveries was that the osmotic pressure depended upon the temperature. The results of his experiments show that the osmotic pressure increases with a rise in temperature. For a rise of 1° *C.* the increase is 1/273 of its value at 0° *C.*, *i.e.* the increase is exactly the same relatively as for the increase in pressure of a gas when its temperature is raised 1° *C.* [*cf.* page 145].

De Vries compared the osmotic pressures of different substances using the same solvent throughout. He was able to show that the osmotic pressure is independent of the kind of dissolved substance, and only depends upon the number of dissolved molecules ; *i.e.* all solutions containing one gramme molecule (1 grm. mol.) of dissolved substance per litre have the same osmotic pressure.

Van't Hoff combined all these laws in one statement : The osmotic pressure exerted by a dissolved substance is exactly the same as the pressure which the same substance would exert if it were present as a gas at the same temperature and occupied the same volume.

To this general rule, however, there is one group of substances which forms an exception—it comprises the electrolytes. It is just this exception, however, which has allowed us to obtain a deeper insight into the nature of solution. This is a point to which we shall return later when discussing electrolytes.

The Brownian Movement. In the last few years quite a new light has been shed upon the phenomenon of osmosis. At first sight these new facts would appear to have little in common with osmosis—we refer to the Brownian movement. In 1827 the English botanist **Brown** had observed very small cylindrical bodies in the sap of plants, which he was examining under a microscope. These particles were continually moving with an absolutely irregular motion. At first it was thought that one was dealing with small living forms, but this phenomenon was eventually observed amongst the particles of solid bodies if they were present in a sufficiently small form. The larger the particles were, the slower did their movements become, but a rise in temperature increased their liveliness. It was then imagined that these movements were due to certain external disturbances, or inequalities in temperature of the liquid ; but if these movements are examined more closely, one is soon convinced that such influences cannot be the primary cause of these movements. For all such agencies would cause all the particles, or at least a group of particles, to move in one and the same direction. Such artificial and intentional disturbances can be produced, and the effect of these is very different from that which actually exists. The most noticeable fact in connexion with the Brownian movement is that **particles which are very near to one another are all moving in different directions, and quite independently of one another.** The second important point is that the motion, so long as there are any particles present, never ceases—neither does it grow less intense even after the lapse of days or weeks.

From such experiments as these one ceases to imagine, but becomes convinced of the fact that we are dealing with traces of a miniature world, which is the basis of our greater world. The phenomenon is also found in gases. If, for example, a single particle of tobacco smoke is observed under a microscope, an entirely similar to-and-fro motion is seen only it is more lively than with liquids. Since it is quite impossible to imagine that these movements can take place on their own accord, the following hypothesis has been put forward, and it has been confirmed again and again by further additional experiments.

If one assumes that the heat of liquids, just like that of gases, is due to the rapid motion of its smallest particles, it follows that small particles of matter, suspended in a liquid, must participate in these movements, for they receive an enormous number of blows every second. But since these particles are several million times heavier than the molecules of a liquid, and the collisions occur very rapidly, and on all sides, it is impossible to observe the effect of a single collision. But the effects in all directions will not be absolutely equal at all times. Small deviations from the mean value will occur more frequently, whereas greater deviations will occur less often as the magnitude of the particles increases. **These departures from the complete equality of the blows in all directions cause the particle to execute this so-called Brownian movement.** We must not imagine that the velocity of the water molecules is diminished after a collision with a heavier particle which is suspended in the liquid as we found when studying the laws of impact, and that these changes in velocity could be observed under a microscope. The true motion and velocity of the solid particles cannot be observed. The number of collisions in a thousandth part of a second is so extraordinarily great and the distance traversed in consequence of these so small, that they are beyond the powers of our perception. Since, however, **the impacts do not follow one another regularly,** when the particle is observed a few seconds later it is found to be in another position. When two such points are joined, as in Fig. 70, we must never believe that the particle has actually traversed this path in moving from one position to another—the real path is infinitely more complicated.

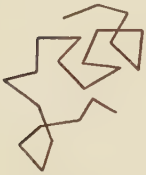


Fig. 70.

Let us now imagine that the solid particles become smaller and smaller. In doing so all the movements will become more lively, the velocities will increase, and the distances which the particles move in the same time will be correspondingly greater; and the small movements which have hitherto escaped our attention because they have been too small to be observed even under a microscope, are now revealed to us. The motion will become more and more complicated, the more quickly we are able to record all its singularities.

Let us now assume that the particles have become so small that we can no longer observe them even with a microscope. Under such circumstances nobody will believe that these motions have ceased because we are no longer able to observe them. On the contrary, they will have become more vigorous and their paths more complicated. Finally, when the particle has become so small, like that of a molecule of sugar dissolved in water, its motion will have become just as vigorous and rapid as is always the case with the particles in a solution. It is certain that we could no longer perceive the motions themselves. When, however, molecules of blue copper sulphate are placed at the bottom of a vessel containing pure water we shall notice that, after a time, the sulphate molecules have moved upwards. **The phenomenon of diffusion only differs from the Brownian motions in that the dissolved particles are very much less in mass.**

If, therefore, a semi-permeable membrane is used to separate a sugar solution on the one side from pure water on the other, we have to imagine that there is a mutual attraction between the sugar particles and the water molecules. The water molecules in the membrane will experience a small pull towards the side nearer the solution. Hence the number of water molecules travelling towards the solution through the membrane will be greater than that of those travelling in the reverse direction. In consequence of this the quantity of water in the solution must increase as we have seen in the example of the cylinder described on p. 132, Fig. 69. This will continue until the excess pressure exerted in the solution is just sufficient to counteract the affinity of the molecules of sugar for the water molecules at the membrane.

Even if we have not yet arrived at a complete explanation of the phenomena of osmosis and of Brownian motion, we observe a very intimate connexion between diffusion, osmotic pressure, and the molecular movements associated with Brownian motion. All these phenomena can be accounted for if it is assumed that the molecules of a liquid are in a state of perpetual motion.

Since the Brownian movements can be observed directly with our eyes, this hypothesis receives a degree of certainty which makes it something which is more than a mere probability.

CHAPTER IV

THE SOLID STATE

WE now come to a class of substances with which, on account of our continual association with them, we are most familiar. Actually, however, we shall soon see that, although we have become familiar with these things, we have no well-founded and intimate knowledge of the structure and mechanism of solid bodies, for it is the solid state of aggregation which has guarded its secrets most obstinately. It must be understood, however, that what has been said already in connexion with elasticity, applies, in general, to all the phenomena associated with solid bodies.

Solid bodies are characterized by the existence of a definite shape in virtue of which we are able to recognize them individually. This implies that the constituent particles of a solid body are fixed relatively to one another, and this of course necessitates a definite form and volume. Strictly speaking, this is not so, for both their shape and volume can be altered by impressed forces. All solid bodies possess a certain degree of form and volume elasticity so that they can transmit both transverse and longitudinal waves. The forces of cohesion are also more marked in solids than in liquids ; these forces are responsible for the definite shape of solid bodies. In crystals these forces imply limits to the shape which the solids can assume. The simple laws which govern the formation of crystals regulate not only the kind and number of atoms which may go to form a crystal, but also define their distances apart, and the orientations of the crystal planes.

Solids fill space and produce the same degree of compactness as do liquids, for in general only small changes in density occur when bodies are melted or solidified.

Since we have learnt to attribute the heat of liquids and gases to a molecular or atomic motion, we cannot imagine that such motions have ceased in solids. Since, however, the particles cannot simply vacate their positions, the hypothesis has been suggested that the particles of a solid body are vibrating about a mean position, as do the prongs of a tuning fork. The particles must be constrained to this position of rest by elastic forces : this is a necessary assumption since the whole body possesses elasticity. The energy is therefore, as with vibrating bodies, continuously alternating between kinetic and elastic forms. Consequently, since any solid body, which is perceptible to our senses, must contain a very large number of molecules, half its total energy must be kinetic, whilst the other half is potential. Even with gases and liquids the energy during the time of impact must be regarded as elastic energy. But since the duration of an impact is so short relatively to the time which elapses between two successive impacts, the portion of the energy which is present as potential energy in fluids is so small, that it is usually neglected.

Crystals. The forces which cause bodies to assume a definite and regular shape, as with crystals, are very important. Since these definite shapes exist in the smallest fragments which we have so far been able to make, we might have reason to hope that these phenomena would lead to a deeper understanding of the structure of solids. Even in our own time this hope has been fulfilled to a greater extent than we ever dared to expect. But many other discoveries had to be made before this new information was forthcoming. We must therefore make a short study of crystals. One of the first facts known about crystals was that the velocity of light in them varied with the direction of propagation. If we consider what has already been stated concerning the velocity with which a disturbance can be propagated through a medium, we see at once that the above fact implies that the properties of a crystal depend upon the direction along which they are examined. This view is confirmed by the fact that most crystals can be split along various cleavage planes. The ease with which mica can be split along one direction into thin sheets is well known, but it is only with great difficulty that mica can be split in any other direction. It is difficult to explain these phenomena if matter has no kind of structure, but consists of an entirely homogeneous mass. On the other hand an easy explanation is forthcoming if we assume that the atoms in a crystal are arranged in regular layers, which cause these differences to exist.

Moreover, there are no phenomena associated with solids which contradict this hypothesis.

The conclusion that liquids and gases are built from the same ‘ bricks ’ as solid bodies now receives more support than would appear at first sight. For if, in this instance, we wished to depart from the idea of molecular structure which we have used in connexion with the other states of aggregation, such a departure would signify that, whenever a transition from one state to another occurred, a quite incomprehensible and thorough metamorphosis of the atom must have taken place. Even if we have the choice between two hypotheses, according to one of which such a complete metamorphosis does happen, whereas the other does not require such a change, there cannot be even a moment’s doubt which explanation is to be chosen, if we wish to form a rational science based on laws. Before we could assume such a far-reaching change, very thorough and convincing proofs would have to be advanced—proofs which would be equivalent to a statement that the alternative view is impossible, because it stands in contradiction to established facts. If we had not got the necessary grounds from which to form an atomic theory to account for the structure of solids, we should still have to refuse any alternative view so long as it was not supported by well-founded facts ; but this state of affairs does not arise, for we do not know a single fact which would contradict the idea of an atomic structure of matter. Now we must endeavour to obtain some ideas concerning the properties of crystals, and in order to do this we must first of all become acquainted with the various systems of crystals. For the extraordinarily large number of crystal forms it is first important to refer back the immense number of possible forms to a few groups, by grouping together those forms with common properties. It was soon found that small and large crystals of the same substance differ from one another only in as much as do two geometrically similar figures. The angles of a crystal, *i.e.* the inclinations of the faces to one another, always remain the same, but the dimensions of a crystal may be either large or small. Everything depends upon the inclination of the faces to one another, and it was only by directing attention to these facts that a system of classifying crystals was made possible.

Through any fixed point inside a crystal certain directions are chosen—they are called the axes of the crystal for they form the axes of a coordinate system to which the characteristics of the crystal can be referred. All crystals, according to the number, position and length of their axis, can be divided into six classes, which are called the **six crystal systems.** They are as follows :

1. The regular or cubic system which has three equal axes mutually perpendicular to one another (Fig. 71, I).

2. The hexagonal system with four axes (II). The three so-called secondary axes are of the same length, they lie in one plane and intersect one another at 60°. The fourth, or principal axis, is perpendicular to this plane and may be of any length.

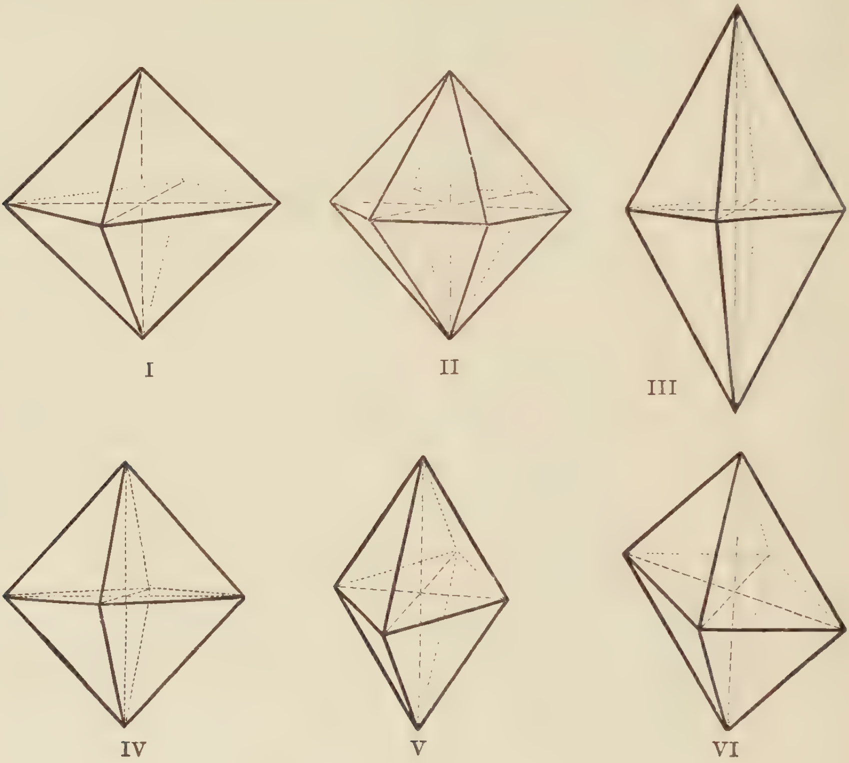


Fig. 71.

3. The tetragonal system has three axes (III) intersecting one another at right angles. Two axes are equal, but the third, or chief axis, can be shorter or longer than these.

4. The rhombic system has three axes (IV) of unequal length but at right angles to one another.

5. The monoclinic system has three unequal axes (V). Two of these intersect at any angle whilst the third axis is perpendicular to the plane containing them.

6. The triclinic system has three unequal axes (VI), intersecting one another at any angle.

By means of these systems all the faces which occur in crystals can be specified according to a simple plan, which is as follows. The position of every face can be fixed definitely when the points are known at which the face intersects the axes. Faces which are parallel to an axis are said to 'intersect it at infinity'. If all these axial distances are doubled, trebled, etc. . . . new planes are specified which are parallel to the first. Hence all parallel planes are characterized by the fact that the ratio of their axial distances is always constant. Since it is only a question of the **direction** of the plane and not of the **magnitude** of its distance from the zero of coordinates, we are thus concerned only with the ratios of these axial intercepts.

Now it is not necessary to have six systems of axes, for we now refer all conceivable planes to a single system. But if the six systems are chosen, it is found that these ratios are always represented by very simple integers (0 and ∞ for parallelism being included), *i.e.* 0, 1, 2, 3, ∞, or their reciprocals, ∞, $\frac{1}{2}$, $\frac{1}{3}$, 0.

In the diagrams shown in Fig. 71 (I–IV), the simplest faces have been indicated. They are represented by the values, 1.1.1. This means that all faces cut the axes at distances which are proportional to the lengths of the axes.

If one of the faces cuts an axis at a distance a while the others are cut at infinity, then this face is parallel to the plane containing these two last axes. If this is true for all the faces of the crystal a very simple figure is formed—for bodies belonging to the first system it is a cube. In the other systems a prismatic figure is formed.

Concerning the physical characteristics of the various systems it may be said that the crystals belonging to them are all homogeneous, *i.e.* all **points** inside a crystal always behave, both qualitatively and quantitatively, in the same way. This is not true when we consider the **directions inside** a crystal—*i.e.* crystals are not necessarily isotropic. The fact that some crystals cleave more easily in one direction than another furnishes us with direct evidence of their anisotropic nature. In addition to this, the conductivity of crystals for heat, light and electricity is different in different directions.

Still more remarkable than these properties are the conclusions which one has been able to make from these differences concerning the arrangement of the atoms in crystals. In general, it is quite evident that the old hypothesis, which regarded matter as a continuum, could not in any way account for these properties which varied with the direction. Even the ideas of elasticity and compressibility presented difficulties. Modern research has so often discovered a large number of differences which depend upon direction, and which can be so easily explained on the assumption that the assemblage of atoms is different along these directions, that the alternative view has been abandoned altogether. Research workers have rather concentrated their attention on discovering the arrangement of the atoms or molecules which will account for these differences. The discovery by **von Laue**, in 1912, that a crystal could be used as a grating to examine the nature of **Röntgen rays** has made it possible for such arrangements to be discovered. Since then very important progress concerning the structure of matter from very tiny units has been made, and the additional task of discovering the nature of the forces which hold these units together has also been undertaken. In many cases it has been possible to obtain trustworthy evidence concerning the arrangement of atoms in a crystal. Generally it is found that the arrangement is one of extraordinary simplicity and regularity. For example, the sodium and chlorine atoms in common salt ($NaCl$) are arranged alternately at the corners of a cube as in Fig. 72. The length of the edge of a unit cube which is also the distance between a sodium atom and its chlorine neighbour can be shown to be $2 \cdot 8 \times 10^{-10}$ cm.

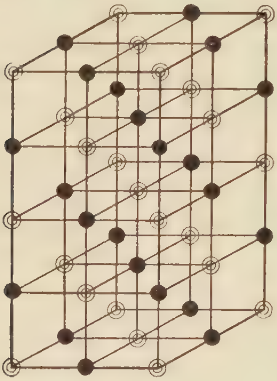


FIG. 72.

It is seen from such an arrangement that every chlorine atom is surrounded by sodium atoms all at the same distance away. We are therefore no longer able to speak of a molecule of sodium chloride, because no single sodium atom is combined exclusively with a single chlorine atom. The cohesion forces are no longer different from the chemical forces by which the atoms of a molecule are held together. Many conclusions have been made concerning the nature of these forces; in general they are of an electric nature, being due to the attraction between the positive and negative charges carried by neighbouring atoms.

In other instances it is necessary to assume that the entire molecule, like that of hydrogen chloride (HCl), remains intact and goes to form a crystal lattice; in such cases it is not the atoms but the molecules themselves which form the units from which the lattice is constructed.

CHAPTER V

TEMPERATURE

1. **THE Comparison of Temperatures.** If we dip our hand into water which has been standing near a fire for some time, we say that the water is hot. When our hand is dipped into melting snow we say that the mixture is cold. Later on, when the water has been removed from the fire, we shall find that it is less hot ; likewise the water from the melted snow ceases to be so cold, when it has been standing in a room for some time. Finally we should find that both quantities of water behave in precisely the same way with respect to our sense of feeling. By further observations we find that heat and cold are different sensations which we experience through our sense of touch and they correspond, moreover, to the different states in which a body may be. This state of a body is called its temperature, and we must now examine the question : Are the observations made with our hands reliable ? No, they are not. Moreover, we have no unit rod with which the temperature interval between two bodies can be measured. We cannot say, with certainty, whether or not two bodies are at the same temperature. If my right hand is placed for a minute into hot water whilst the left dips into melting snow, and both are afterwards plunged into a third vessel containing water at room temperature, then according to the sensation in my right hand, this water is cold, whereas according to my left it is warm. Our eyes tell us quite clearly that both hands have been in the same water whilst our sense of touch has experienced two different impressions. Now, since we cannot ascribe to one quantity of water two temperatures at the same time, we are forced to conclude that our sensation of touch is not a reliable one wherewith temperatures can be compared. We have to look around for other means to measure temperature ; in these the sense of touch will be excluded if possible, and we shall seek a method which will give us quantitative information concerning temperatures.

2. **The Measurement of Temperature by a Thermometer.** Such a method is furnished by the expansion of bodies. We can easily convince ourselves that, in general, all bodies whether solid, liquid, or gaseous, expand when heated. Telegraph wires sag more in summer, than in winter. If a flask having a long narrow neck is filled with water, and then placed near a fire, the level of the water soon begins to rise. Or if a flask merely filled with air is heated, and then allowed to cool whilst its neck is immersed in water, the water will rise in the flask, showing that some air had been expelled during the process of heating.

With the aid of these preliminary temperature indicators we arrive at the following conclusions. All bodies which have been in contact for a sufficiently long time acquire the same temperature. The equality of the temperature is determined by the fact that any test body expands by the same amount irrespective of the particular object with which it is in contact. We therefore have a very easy and useful means of imparting the same temperature to two bodies. We can conclude that the thermal expansion of the test body is conditioned by its assuming the same temperature as that of the body under investigation, when it is brought near to or in contact with that body, and that its expansion reaches a limiting value when it has completely attained this temperature.

It is of course necessary to discover some temperature which is invariable. It has long been known that a mixture of melting snow, or ice, and pure water always has the same temperature. No matter where the experiment was made or what method was employed to measure the temperature, it was always constant, *i.e.* the dilation of the body used to measure the temperature was always constant whether the test was made in summer, or in winter, here, or in America. This temperature is referred to as that of melting ice, and it is employed as the zero temperature to which all other temperatures are referred. But another temperature needs to be defined before we can say that some other temperature is one degree (or ten) higher, or lower, than that of melting ice, which we have chosen to be our zero. This could be defined quite arbitrarily. For example, we might say that a certain wire has become 'one degree' hotter because its length has increased by one millimetre. If this were done, then all other measurements, which had to be compared with ours, would have to be made with this same wire or else all other thermometers would have to be compared with ours. This would be very inconvenient and practically impossible owing to the fact that temperature measurements are constantly being made everywhere. We therefore have to seek some other temperature which is constant, and either choose the difference between these two temperatures as the unit, or else define this as an arbitrary multiple of the unit of temperature. **Newton** knew that the temperature of the human body is very constant and called this a temperature of one hundred ; this scale is very nearly the same as that devised by **Fahrenheit** which is still used almost exclusively in this country. Ultimately, however, it was discovered that the temperature of the body was not constant enough, so that it became necessary to define another fixed point in terms of some property of water. Under constant pressure water always boils at the same temperature. This temperature is considerably higher than that of melting ice, so that the difference is too large to be used as a unit ; it is therefore divided into one hundred equal parts called **degrees.**

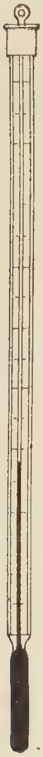


Fig. 73.

It must be mentioned, however, that although a temperature can be measured with several different types of thermometer, in practice it is found that some are not so convenient to use as others. The type most generally used is the so-called mercury-in-glass thermometer. This contains a small mass of mercury the expansion of which serves to indicate changes of temperature. The mercury is contained in a small glass vessel which may be either cylindrical or spherical in shape. This bulb is joined to a long capillary glass tube which may vary from 10 to 50 cm. or more in length. When the mercury expands, owing to a change in temperature, it rises in the capillary tube. Since this is very small in diameter any small change in the volume of the mercury is easily measured, since the mercury thread moves a considerable distance along the tube. A thermometer having been made it is placed in boiling water, or better in the steam from boiling water. On the level near to the end of the mercury column a scratch is made on the glass and it is marked 100. The thermometer is then placed in melting snow and the final position of the mercury thread is marked by another scratch—the zero of the thermometer. The distance between these two marks is divided into one hundred equal parts—it is assumed that the bore of the tube is uniform. Each such division is defined as one degree on the **Celsius** or **Centigrade** scale of temperature, and is written 1° *C*. We do not propose to discuss the art of measuring temperature in detail, and will only mention the fact that the gas thermometers are the precision instruments—they depend upon the fact that the so-called perfect gases expand when heated. All other thermometers can be compared with such standard thermometers at such places as the Reichanstalt, Berlin, or the National Physical Laboratory, Teddington.

3. **The Expansion of Solid Bodies.** Let us now investigate with the aid of a thermometer the expansion which occurs when solid bodies are heated. If we take rods or wires of different materials we shall find, what is almost self-evident, that the change in length is proportional to the length of material itself. For since we assume conditions to be the same, and cause the wire to be at the same temperature everywhere, each centimetre of wire must expand by equal amounts. The total change in length is equal to the sum of the changes occurring at each point in the wire. Hence if we wish to compare the behaviour of different bodies with respect to temperature, equally long pieces must always be compared. We therefore try to determine the change per unit length. If the wire or rod is of length l_0 at 0° and l_t at $t°$, then the change per unit length is $\frac{l_t - l_0}{l_0}$. This change has been produced by raising the temperature $t°$. If we now assume that the change is proportional to the temperature difference, then the expansion per unit length per unit rise in temperature is

$$a = \frac{l - l_0}{l_0 t} \qquad . \quad . \quad . \quad . \quad . \qquad (64)$$

The value a is called the coefficient of linear expansion of the given substance. For solid bodies a is generally a few millionth parts of unity. The following table gives a few values of a for some common substances.

TABLE 10

Coefficients of Expansion × 10^6.

Substance.	a.	Substance.	a.
Aluminium	23	Glass	8–10
Lead	29	Porcelain	3–4
Iron	11	Ice	50
Copper	16	Fused quartz	0·7
Brass	19	Oak :	
Nickel-steel	1·5	Across the grain	54
Platinum	9	Along the grain	5

The equation (64) enables us to calculate the length of a body at any temperature $t°$ *C.* when its length at 0° is known. We have

$$l_t = l_0 \, (1 + at) \qquad . \quad . \quad . \quad . \qquad (65)$$

This equation signifies, when it is considered as being strictly correct, that the change in length is proportional to the increase in temperature. Critical researches have shown that this is not so. One has to remember, therefore, that the coefficient of expansion is itself a function of the temperature. For considerable changes of temperature this fact must be incorporated in an equation, when exact calculations have to be made.

When we have to consider the expansion in more directions than one we must remember that an area expands along its whole length and breadth, whilst bodies expand in all three perpendicular directions. For homogeneous isotropic bodies the coefficients of expansion are the same in all directions, but for other bodies this is not so. In wood, for instance, the coefficient is much less along the grain than in a direction perpendicular to it. Also in crystals with unequal axes, the coefficients generally vary with the particular direction along which they are measured.

4. **The Expansion of Liquids.** One of the most remarkable facts observed in the expansion of liquids is that the density of water is a maximum at 4° *C.*—otherwise nothing very important is to be noted. In other words, when water is heated from 0° to 4° it does not expand, but contracts. At 8° the volume is practically the same as at 0°. This phenomenon is explained on the hypothesis that the molecules of a liquid form clusters before passing into the solid state. In connexion with liquids, one only speaks about a coefficient of cubical expansion, *i.e.* of all three dimensions. This is generally greater than the corresponding one for solids. The coefficient of cubical expansion is defined by means of an equation which resembles the one we have already used for linear expansions, viz.

$$\frac{V_t - V_0}{V_0 t} = a$$

or

$$V_t = V_0\ (1 + at) \quad . \quad . \quad . \quad . \qquad (66)$$

A few such coefficients for liquids are given below.

TABLE 11

Coefficients of Cubical Expansion × 10^6.

Substance.	*a.*	Substance.	*a.*
Mercury	180	Alcohol	1100
Petroleum	990	Ether	1600

In the determination of the expansion of liquids we must remember that the vessels containing them also expand. What we directly observe is the difference of the two cubical expansion coefficients, *i.e.* of the liquid and of the vessel.

5. **The Expansion of Gases.** The pressure under which a gas is investigated exerts a considerable influence upon its expansion. By changing the pressure the volume of a gas can be altered at will. It is therefore necessary to use two coefficients of expansion when dealing with gases.

(1) The **coefficient of increase in volume** which is involved when the temperature increases, but the pressure remains constant.

(2) The **coefficient of increase in pressure,** deduced from the excess pressure which must be applied in order to keep the volume of a gas constant after its temperature has been raised. One could imagine that the gas first expands under constant pressure, and that it is subsequently compressed according to Boyle's Law, the temperature remaining constant. When the increase in pressure is such that the original volume has been reached the coefficient of increase in pressure at constant volume can be calculated. At first it is necessary to draw a distinction between these two coefficients. We have

$$\frac{V_t - V_0}{V_0 t} = a_v, \text{ or } V_t = V_0 (1 + a_v t), \text{ when } p = \text{const.} \quad . \quad . \quad (67)$$

and

$$\frac{p_t - p_0}{p_0 t} = a_p, \text{ or } p_t = p_0 (1 + a_p t), \text{ when } V = \text{const.} \quad . \quad . \quad (68)$$

When, in 1802, **Gay-Lussac** determined the coefficients of expansion for air, and some other gases, always taking care that the volume was measured under the same pressure, he obtained the striking result that the coefficient, a_v, is the same for all gases. Later and more exact determinations have shown that **Gay-Lussac's** law is only a first approximation to the real facts: like Boyle's Law, it is almost true for the so-called permanent gases, whilst for others it becomes more exact the lower the pressure and higher the temperature of the gas, *i.e.* it becomes more exact when the gas is further removed from the state when it will liquefy. The value of a_v according to the best determinations is 1/273·2 or 0·003660, *i.e.* all the permanent gases expand by 1/273·2 of their volume at 0° *C.*, when the temperature is raised 1° *C.* For perfect gases this value would be independent of the pressure and temperature; for actual gases, however, this value is more closely approached the further they are removed from the point at which they become liquid.

6. **The Effect of Heat on Gases at Constant Volume.** The relationship between the temperature and pressure of a gas can be obtained without making any further experiments, by using the laws of Boyle and Gay-Lussac. Let us assume that, at first, we have a gas at atmospheric pressure p_0. Let its volume at 0° *C.* be V_0. Then the volume V_t at some other temperature t is given by the above equation, $V_t = V_0 (1 + a_v t)$. If the gas is now compressed at constant temperature, the product of the volume and pressure is, according to Boyle's Law, constant. The pressure of the gas is p_0 when its volume is V_t. If p_t is the pressure when the volume V_0 has again been reached, then

$$p_t V_0 = p_0 V_t = p_0 V_0 (1 + a_v t).$$

Whence it follows that

$$p_t = p_0 (1 + a_v t)$$

or

$$\frac{p_t - p_0}{p_0 t} = a_v,$$

and this fraction has been defined in (68) as the pressure coefficient at constant volume, and denoted by a_p. It therefore follows that for ideal gases $a_p = a_v$ and that these values are independent of the temperature and pressure conditions under which they are measured. For all actual gases a_p and a_v are different to an extent which depends upon how far the gases obey the above laws. These conclusions have been completely confirmed by experiment. It is very easy to ascertain the increase in pressure which arises when a gas is heated under constant volume conditions so that the pressure coefficient a_p is determined directly. The following table indicates the magnitude of the divergences which occur.

TABLE 12

Gas.	a_p.	a_v.
Helium	—	0·003665
Hydrogen	0·003660	0·003662
Nitrogen	0·00367	0·00367
Oxygen	0·00367	0·00367
Air	0·00367	0·00367
Carbon dioxide	0·00370	0·00370
Sulphur dioxide	0·00390	0·00384

Accordingly in future, when we deal with perfect gases, we shall no longer make a distinction between a_p and a_v but write $a_p = a_v = a$.

Now according to equation (56) there must be a definite mean velocity with which all the molecules of a gas move, if we wish to attribute the pressure of a gas to molecular collisions. Since we wish to do this for all pressures, we must investigate how, according to this assumption, the pressure of a gas can vary. We will assume that the volume remains constant, and that the temperature changes according to equation (68). Now since neither the number of molecules present,* nor their mass, can alter, it follows that a change in pressure at constant volume can only be due to a change in the velocity of the molecules themselves. Hence to every temperature there must correspond a definite mean velocity of the molecules. Let this be v_0 at 0°, and v_t for $t°$. Substituting these values in equation (56) [*cf.* page 120], we have

$$p_0 = \frac{Nmv_0^2}{3V}, \text{ and } p_t = \frac{Nmv_t^2}{3V}.$$

But since $p_t = p_0(1 + at)$, it follows, by eliminating p_t and p_0 from the above equations, that

$$v_t^2 = v_0^2(1 + at) \quad . \quad . \quad . \quad . \quad (69)$$

From this equation the temperature coefficient for the mean square velocity of the molecules can be determined by solving this equation for a. We have

$$a = \frac{v_t^2 - v_0^2}{v_0^2 t} \quad . \quad . \quad . \quad . \quad . \quad (70)$$

This coefficient a therefore tells us the fractional increase in the mean square molecular velocity when the temperature is raised 1° *C.*, the volume remaining constant. Numerically it is equal to 1/273·2.

The energy of a body is determined by the square of its velocity, so that we may consider the above equation as telling us the fractional change in the mean energy due to the motion of the particles, when the temperature increases 1° *C.* These remarks may perhaps be made a little clearer by multiplying both the numerator and denominator of equation (70) by $m/2$ for then the terms which occur on the right-hand side of the equation each represents a certain amount of energy associated with a molecule.

---

* We exclude the case of dissociation in which there is an increase in the *number* of molecules as the temperature increases.

7. **The Absolute Zero of Temperature.** Most gases remain as gases when their temperature is below 0° *C.*, so that the value of t in equations (65) to (70) may be negative. The expression $(1 + at)$ will then be less than unity. Ultimately its value will become zero, when $at = -1$, *i.e.* $t = -\frac{1}{a} = -273{\cdot}2°$. Under these circumstances v_t^2 would be zero according to equation (69). The molecules would no longer be moving, so that it would be impossible for them to exert a pressure on the walls of the containing vessel, *i.e.* $p_t = 0$. Also the volume V_t would be 0, a state of affairs which it is very difficult for us to visualize. But it has already been mentioned that, in our deduction of the above formula, we have neglected the space occupied by the molecules themselves, and assumed that the whole of the space, without any reservation, was at the disposal of the molecules. Actually it is only the space to which the molecules have resource for their movements that becomes zero—a result which is quite easy to comprehend.

It would therefore be much more logical, not to measure temperatures on the basis of that of water being zero, but to select a zero which is 273·2° below that of melting ice. This is called the **absolute zero of temperature** because at that point thermal motions would cease entirely. The carrying out of this idea has shown that all our equations become more simple, so that they can be expressed in a more lucid manner. If we therefore again call the temperature difference between melting ice and boiling water one hundred degrees of temperature, it follows that all temperatures expressed on the absolute scale will be numerically 273·2° higher than the corresponding temperatures on the Centigrade scale. Where we have $t = 0$ we shall therefore have to write $T_0 = 273{\cdot}2 = \frac{1}{a}$ when using the new scale of temperature ; where we now have t, we shall get $T = 273{\cdot}2 + t$. Also wherever the expression $(1 + at)$ occurs the old value of t will have to be replaced by its new one. But since

$$t = T - T_0, \text{ and } a = 1/T_0,$$

we have

$$1 + at = 1 + \frac{T - T_0}{T_0} = \frac{T}{T_0}, \qquad . \quad . \quad . \quad (71)$$

so that equations (67), (68) and (69) become

$$\left.\begin{aligned} V_t : V_0 &= T : T_0 \ (p = \text{const.}) \\ p_t : p_0 &= T : T_0 \ (V = \text{const.}) \\ v_t^2 : v_0^2 &= T : T_0 \end{aligned}\right\} \qquad . \quad . \quad . \quad (72)$$

We shall soon have an opportunity of discussing the behaviour of bodies at the absolute zero.*

8. **The General Form of Gas Equation.** With the aid of the laws of **Boyle** and **Gay-Lussac** the volume of a gas can be calculated when, (1), the temperature is raised, and (2), the pressure has any desired value. Since the two laws are applied one after the other, this problem may be considered as being solved. But we can combine the result in one formula as **Clapeyron** first did (1799–1864), and obtain an equation in which

---

* Temperatures on the absolute scale are often denoted thus, 452° *K*. The *K* is used in honour of Kelvin, who first showed that the absolute scale was thermodynamically sound. [*Trans.*]

the pressure, volume and temperature occur, so that when two are known the third is fixed. This equation is the general gas equation. It is obtained as follows :

Let us assume that we have a perfect gas at 0° C., and under 760 mm. pressure, in a volume V_0. We shall first heat the gas to a temperature $t°$, and then compress it to a pressure P. We require to find the volume of the gas under these new conditions.

We will imagine that this final state has been reached in two stages—the order is immaterial. Thus the temperature may first be raised from 0° to $t°$ under constant pressure P_0 (atmospheric pressure). The new volume is given by Gay-Lussac's law—it is $V_t = V_0(1 + at)$. But this volume is not the one which is required, for the gas is still at atmospheric pressure although its temperature has attained the desired value. We therefore maintain this temperature and alter the volume by subjecting the gas to a pressure P. Now this variation will be governed by Boyle's Law—the product of the pressure and volume is constant. When the volume is V_t and the pressure P_0 the product is V_tP_0. If the new volume is V and the pressure to which it is subjecte d P we therefore have $VP = V_tP_0$. Replacing V_t, which is not known initially, in terms of V_0 we have

$$VP = V_0P_0(1 + at) \quad . \quad . \quad . \quad . \quad (73)$$

This is the general gas equation. If the temperature is calculated on the absolute scale we must put $1 + at = T/T_0$, so that we obtain

$$VP = V_0P_0T/T_0,$$

an equation which can be more conveniently written as

$$\frac{VP}{T} = \frac{V_0P_0}{T_0} \quad . \quad . \quad . \quad . \quad . \quad (74)$$

On the right-hand side of this equation we find a quantity which is invariable ; consequently the left-hand side is constant. This fact shows that however we change two of the three variables V, P and T, the value of the third is such that PV/T remains unaltered.

Now let us calculate, if possible, the value of this invariable quantity. P_0 is taken as being the standard pressure due to one atmosphere so that its value in absolute units is $1{\cdot}013 \times 10^6$ dynes/cm.². T_0 is the temperature of melting ice, viz. 273·2° K. In addition, we have V_0, the volume of the mass of gas to which the equation refers. Consequently when different quantities of gas or even different kinds of gas are to be considered, P_0 and T_0 are defined whilst V_0 is still arbitrary. **But if such quantities of gas are considered so that they always have the same volume at 0° C. and under one atmosphere pressure, the volume V_0 is the same for all substances,** and consequently the right-hand side of the above expression is **the same for all gases.** If therefore it is agreed that only such quantities of gas shall be compared which have the same volume at $N.T.P.$, then for all gases, under whatsoever conditions of temperature and pressure they may be, the value of PV/T is always constant.

It has been agreed to consider such quantities of gas that their mass, expressed in grammes, is equal to their molecular weight ; thus, 2 grm. of hydrogen, 32 grm. of oxygen, and 4 grm. of helium, etc. This quantity is called a gram-molecule. Since they all contain the same number of molecules they have, according to **Avogadro's** Law, the same volume. By means of a balance it has been found that a litre of nitrogen at normal temperature and pressure has a mass of 1·25 grm., so that a gram-molecule (or 28 grm. of nitrogen) occupies 22·4 litres. Hence the volume of a gram-molecule of any gas at *N.T.P.* is 22·4 litres. If such a quantity of gas is used for the purposes of calculation, the expression on the right-hand side of equation (74), in absolute units, is

$$\frac{P_0V_0}{T_0} = 1{\cdot}013 \times 10^6 \times 22{,}400/273{\cdot}2 = 8{\cdot}3 \times 10^7.$$

This value is a general one common to all gases. If this constant is denoted by R, as is usual, the general equation for gases can be written

$$PV = RT \quad . \quad . \quad . \quad . \quad . \quad (75)$$

Since the expression on the left-hand side represents a quantity of work [*cf.* page 126], so must the right-hand side also be equal to the same amount of work. By raising the temperature 1° *C.*, PV must increase in value by an amount R. The value of this for all gases is $8{\cdot}3 \times 10^7$ ergs, or 8·3 Joules, per 1° *C.* It is therefore 1·98 cals., so that, as a first approximation, we may say, $R = 2$ cals.

We are now in a position to obtain a clear idea of what the word temperature implies. For the same gas, it has been shown above that the temperature is proportional to the kinetic energy of the molecules. Ought this relationship to be extended to the case in which we deal with different gases ?

We have shown the equation $PV = RT$ to be true in general. Previously we have obtained, as the fundamental equation of the kinetic theory of gases,

$$PV = Nmv^2/3,$$

so that

$$RT = Nmv^2/3.$$

For any other gas we should have similarly

$$RT_1 = Nm_1v_1^2/3,$$

since we always work with a gram-molecule of gas—a quantity in which the number of molecules is always the same, and equal to N. The condition, therefore, that two gases should have the same temperature, $T = T_1$, is that

$$mv^2 = m_1v_1^2 \quad . \quad . \quad . \quad . \quad . \quad (76)$$

Gases have equal temperatures when, with respect to their rectilinear motion, their molecules possess the same kinetic energy. Since, at different temperatures, the kinetic energy of all gases is proportional to their temperature, the result is still true for all gases and temperatures, *i.e.* the temperatures of two gases are equal when the mean molecular kinetic energy due to rectilinear motion is the same for each gas.

Attention may now be called to the fact that the equation $PV = RT$ is only strictly true for perfect gases. Actual gases only obey this law in so far as they obey the laws of Boyle and Gay-Lussac—the so-called permanent gases behave in this manner. For other gases, the higher their temperature, and the lower their density, the more closely do they approximate to these ideal conditions. The deviations from the ideal law are represented very closely by the equation of **Van der Waals** [*cf.* page 124], which we must now write in the form

$$\left(P + \frac{a}{V^2}\right)(V - b) = RT \quad . \quad . \quad (77)$$

Reference Library

CHAPTER VI

SPECIFIC HEAT

1. THE IDEA, UNIT, AND DETERMINATION OF SPECIFIC HEATS

THE temperature of a body alone is not sufficient to define its thermal condition. This fact can easily be shown to be true. Let us assume that we have a beaker containing 1 kgrm. of water at 0° *C.* In a second vessel let us heat a similar quantity of water to 100° *C.* and then mix the two quantities together. It will be supposed that the operation has been so carefully performed that all the water at 100° has been transferred without losing anything in mass, or of its heat content. The mixture is stirred, until a thermometer which has been inserted in the liquid shows a steady temperature. This steady condition will be attained when the thermometer registers a temperature in the region of 50° *C.* Now let a second experiment be performed in which 1 kgrm. of lead is heated to 100° *C.*, and then mixed with the same mass of water at 0° *C.* In order to ensure a rapid transference of heat from the lead to the water, it is advisable to use the lead in the form of small lead shot. Under these circumstances we shall find that the final temperature indicated by the thermometer is about 3° *C.* The kilogramme of lead is only able to raise 1 kgrm. of water about 3°, whilst the lead itself cools through $100 - 3 = 97°$, as compared with a kilogramme of water which raises the temperature of the same mass of water 50°, whilst itself cooling through an equal interval of temperature.

Similarly, we can show that when 1 kgrm. of lead shot is placed in a beaker and warmed by a flame supplying a regular amount of heat, it requires, say, five minutes for the temperature to be brought from 0° *C.* to 100° *C.* If a similar mass of water is placed in a beaker and heated under similar conditions, 2½ hours will elapse before the water has attained a temperature of 100° *C.* assuming of course that no heat is lost. From such observations as these we arrive at the conclusion that different substances require different amounts of heat in order for their temperature to be raised by equal amounts.

It is at once necessary for us to decide what shall be the unit quantity of heat. Originally this unit was defined as the amount of heat required to raise one gramme of water 1° *C.* This quantity is called a calorie. Nowadays two calories are in general use ; they are the small, or gram calorie, and the large, or kilogram calorie. The particular one which is used will depend upon whether one gram or one kilogram of water is used in defining the unit. But later on it was found that the unit of heat was not independent of the temperature as had originally been supposed. The amount of heat required to raise water from 0° to 1° is not the same as that required to raise it from 99° to 100°. It is therefore necessary to define our unit in stricter terms. **Regnault** assumed that the heating took place between 0° *C.* and 1° *C.* and defined his unit accordingly.

Other workers have employed the mean calorie which is defined as one-hundredth part of the heat necessary to bring water from 0° to 100° *C.*

In recent years, owing to the fact that experiments are generally made at room temperatures of about 15° the unit of heat which has been chosen is the amount of heat required to raise 1 grm. of water from 14·5° to 15·5° *C*. This 15° calorie is almost exactly equal to the mean calorie but is somewhat greater than Regnault's calorie; in fact it is 1·008 times as big.*

When, moreover, we assume that the quantity of heat necessary to heat a mass of water is greater as the mass of the water is increased, we are really supposing that the quantity of heat necessary is proportional to the mass of the substance, when the kind of substance does not change. We are justified in making this assumption, because with pure water, as with all pure unitary substances, we have found that its properties are determined solely by the kind and quantity of material used. An experiment designed to show that the capacity for heat of two equal amounts of water is the same, would confirm this fact. Then we can recognize all substances from the quantity of heat necessary to raise 1 grm. through 1° *C*. This number is called the **thermal capacity** of the substance. The **specific heat** of a substance is its thermal capacity compared with that of water at some standard temperature. The specific heat of any particular substance is, like its specific gravity, constant; but like the specific gravity it depends upon the temperature, as in the case of water, to which reference has already been made.

The Determination of Specific Heats. The fact that a body has a specific heat s means that in order to raise the temperature of a mass m from $t°_1$ to $t°_2$, $ms\ (t_2 - t_1)$ calories are necessary, and conversely, when the body cools from $t_2°$ to $t_1°$ a similar amount of heat is given out. In order to measure this number of calories, they are transferred, without loss or gain, to a known mass of water, and the rise in temperature of the water is measured. If the initial temperature of the water is t'_1 whilst t'_2 is its final temperature, the heat received by the water is $M\ (t'_2 - t'_1)$ if M is the mass of the water. If the experiment has been carried out so carefully that all the heat has been given to the water, then these two quantities are equal.

Generally the experiment is made in such a way that a known mass of the material (temp. $= t_2$) whose specific heat is required is added to a known quantity of water (temp. t_1). The final temperature is the same for both bodies and it is observed by means of a thermometer. If this is t, we have $t_1' = t_2' = t$, whence $ms(t_2 - t) = M(t - t_1)$. The value of s is therefore given by

$$s = \frac{M(t - t_1)}{m(t_2 - t)} \quad . \quad . \quad . \quad . \quad (78)$$

In this way the specific heats of both solids and liquids can be found, providing that there is no chemical reaction between the substance under investigation and the calorimetric liquid.

This method of mixtures was used in England about 1760 by **Black** and it led him to form the idea of specific heats. In deriving the above equation we have made use of the idea that the quantity of heat which passes from the hot body is equal to that gained by the cold body, since it is 'obvious' that heat can only be transferred from one body to the other. But we are only permitted to do this, if heat is an invariable entity. In fact, at this time, heat was considered to be a substance, an idea which, either consciously or unconsciously, was substantiated by that which assumed fire to be an element. Although this is an erroneous idea yet, in this particular instance, it was not contradicted by experience, since even the hypothesis which asserts that heat is a form of energy, justifies our use of the equality sign in the above equation. For, in particular, since no heat is changed into any other form of energy when the substances are mixed, the fact that the two quantities of heat are equal, is an immediate consequence of the principle of the conservation of energy.

---

* Strictly speaking, the calorie is best defined as one-fifth of the heat required to raise the temperature of one gramme of water from 15° to 20° *C*., this being the smallest interval of temperature which can be measured with a precision demanded by modern science and industry. [*Trans.*]

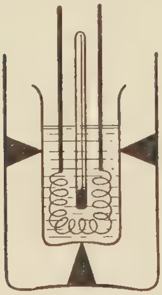


Fig. 74.

Another more recent method, which is only applicable to liquids, is as follows. A thin wire in the form of a spiral is immersed in the liquid and heated electrically. The wire, under the same conditions, is then immersed in an equal mass of water for an equal interval of time. The ratio of the change in temperature in the two cases is equal to the inverse ratio of the specific heats [if we neglect the water-equivalent of the calorimeter].

The following table gives the specific heats of some substances and indicates the order of magnitude of this quantity.

Table 13

Specific Heats of Solids and Liquids.

Water	1·0	Glass	0·19
Lead	0·03	Common salt	0·21
Iron	0·11	Alcohol	0·58
Gold	0·03	Ether	0·56
Copper	0·09	Petroleum	0·51
Carbon	0·2	Mercury	0·033

The table shows that water has an extraordinarily high specific heat; amongst all other known substances there are only a few which have a specific heat approaching that of water in magnitude. In Nature this is a fact of very great importance. At the equator large masses of water are continually receiving heat. In consequence of this, the surface of the ocean rises, and water flows to the cooler regions of the earth. In so doing this water is cooled and imparts its heat to its surroundings. The more dense and cooler water sinks to great depths, and flows back to the equator where it is again warmed (Gulf-stream). These two processes tend to equalize the distribution of temperature over the globe so that mankind can exist in regions where it might otherwise be too hot or too cold. On account of its great specific heat water is the liquid most suitable for this purpose.

Strictly speaking, one ought also to consider the fact that, when a body is heated, it expands, and does work against the extremely applied pressure. The amount of this work, relative to that required to heat the body, is so small that it can be neglected.

2. The Specific Heats of Gases

It is otherwise with gases. With their relatively small mass per unit volume, and large expansion on heating, one cannot neglect the work done on expansion in comparison with the energy required to raise their temperature. It is found convenient to consider the effect of heat on gases under two sets of conditions. First, the pressure can be increased so that the volume remains constant, the temperature only increasing. Secondly, the pressure can remain constant. In this instance the gas expands in consequence of the heat received and performs work against the external pressure. We therefore have to make a distinction between the two specific heats of a gas—we have the specific heat at constant volume C_v and the specific heat at constant pressure C_p. The work done against the external pressure is called the external work, whilst that used in heating the gas is called the internal work.

We have already seen that two gases have equal temperatures when the kinetic energy due to the rectilinear motion of the individual molecules has the same mean value in each case. Since this is true at all temperatures, it follows that the kinetic energy of each molecule must increase by the same amount when a gas is heated. This must also be true when the same number of molecules of different gases are similarly heated. *A priori* we might therefore expect to obtain a very important result when, instead of considering unit mass, we make our calculations with reference to a gram-molecule or gram-atom. We now propose to do this and will let C_v and C_p denote the two specific heats under these conditions.

The volume energy of a gram-molecule of any gas is $PV = RT$, where R is a universal constant for all gases. If the gas is heated at constant pressure through 1° C. the volume will increase and becomes $(V + \Delta V)$, when $P\ (V + \Delta V) = R\ (T + 1)$. Subtracting these two equations we have $P\Delta V = R$, an expression which measures the external work done when the temperature is raised 1° C. **It is the same for all gases and equal 8·3 Joules or 1·98 units of heat;** in short we may say 2 cals.

When, however, the gas is heated at constant volume, *i.e.* the expansion of the gas is reduced to zero because the fixed sides of the vessel in which it is heated do not move, we have $\Delta V = 0$, and no external work is performed against external forces. Moreover, the internal condition of the gas will have been changed by the same amount as in the first case. Hence it follows that, **for all gases**, the heating of a gram-molecule at constant pressure requires 2 cals. more heat than when the experiment is performed under conditions of constant volume. The equation

$$C_p - C_v = 2 \text{ cals.} \qquad . \quad . \quad . \quad . \quad (79)$$

is true for all gases.

If the internal work done, when a gas is heated, merely consisted of increasing the rectilinear motion of the molecules, then the process of warming a gas would be quite clear, for we have found that, at all temperatures, the kinetic energy $E_k = \frac{3}{2}E_v$, so that the increase in the kinetic energy $= \frac{3R}{2} = 3/2$ times the increase in the intrinsic energy, *i.e.* it is approximately 3 cals. This is a measure of the work done in heating a gram-molecule at constant volume; since 2 cals. more are required to heat the same quantity of gas at constant pressure, the ratio of the two specific heats is

$$\frac{C_p}{C_v} = \frac{5}{3} = 1{\cdot}66 \quad . \quad . \quad . \quad . \quad . \quad (80)$$

Is it possible for the molecules of a gas to possess any other kinetic energy than that due to their rectilinear movement ? This would undoubtedly happen if the molecules had an additional rotatory motion. Now most of the common gases have two atoms in a molecule, and we can only conceive these as being joined together to form a sort of dumb-bell. Amongst the great number of collisions which occur, it cannot be otherwise than that sometimes only one of the two atoms will be hit, as is depicted in Fig. 75. This will cause the whole to rotate—in fact both atoms will revolve around their common centre of gravity. This energy of rotation will be obtained at the expense of that due to the linear movements of the molecules. The collisions on the walls of the vessel will therefore become less numerous since collisions on the walls do not occur in virtue of the rotational energy of the molecules. A molecule can keep on rotating without making any appreciable contribution to the pressure on the walls. Similarly a thermometer which is placed inside the gas is not affected by the rotating molecules, but only by those which are moving in a straight line. In other words, the molecular energy due to rotation, as such, does not increase the temperature of the gas ; that will only happen by chance, when a collision has changed the motion into a linear one. When therefore the temperature of a gas has to be raised one has to supply an extra amount of energy which is equal to the gain in rotational energy. It is therefore to be expected that the specific heat will be greater when the number of atoms per molecule increases. In molecules which consist of several atoms, four or more, a great variety of types of intermolecular motion is to be expected ; first, various movements in which the particles approach and recede from each other in a vibratory manner ; then there is also another variety which depends upon whether the rotation takes place around one axis or another, so that the relations become quite indefinite. For the more simple types of molecule consisting of two or three atoms, **L. Boltzmann,** in his exhaustive treatise on the kinetic theory, has obtained results which have been confirmed many times by experiment. Since the pressure on each wall of a cubical vessel is the same, and since this pressure, according to the fundamental equation of the kinetic theory, is intimately associated with the kinetic energy of the molecules which travel to and fro between the walls, it follows that if we resolve the whole energy due to rectilinear motion into three components along directions parallel to the edges of the cube, these three components must be equal. Since the total energy is equal to $3RT/2$ the amount of energy associated with each component is RT/2. If the temperature is raised one degree, then the energy of each component will increase by an amount $R/2$.

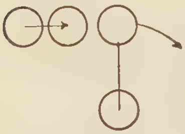


Fig. 75.

Boltzmann assumed that the atoms themselves are so tiny that single atoms will only rotate on very rare occasions, and even when they do rotate the energy will be so small that it does not affect our present calculations.

For **monatomic gases,** therefore, the only kinetic energy present is that due to the rectilinear motions of the molecules. Hence, in raising the temperature of a monatomic gas, when there is no change in volume, it is only the kinetic energy $C_v = 3R/2 = 3$ cals. which has to be supplied. But when the volume increases under constant pressure external work has to be done in addition. Hence the specific heats of a monatomic gas are, according to theory, $C_p = 5$ cals. and $C_v = 3$ cals. The ratio $C_p : C_v = 5 : 3 = 1{\cdot}66$. These results were published by **Clausius, Maxwell,** and **Boltzmann** at about the same time and before any monatomic gas was known, for at that time the existence of such a gas had not been suspected. Soon afterwards, however, **Kundt and Warburg** were able to show that for mercury vapour, which on chemical grounds was known to be monatomic, the ratio of the specific heats was 1·66. Twenty years later, **Ramsay** and **Rayleigh** found in the atmosphere a whole series of the so-called inert gases, which did not enter into chemical combination with any other substances. These gases were monatomic. For helium and argon the specific heats themselves have been measured, and the results are recorded in the following table.

TABLE 14

Gas.	Mol. Wt.	C_p.	C_v.	C_p/C_v.	$C_p - C_v$.
Helium	4	5·0	3·0	1·66	2·0
Argon	39·9	5·07	3·07	1·66	2·0

The table shows that these numbers offer a striking confirmation of a theory which had been developed before the existence of these gases had even been suspected. The external work is 2 cals. more than the increase in internal energy, viz. 3 cals., whilst their sum is exactly equal to the value of C_p as found experimentally.

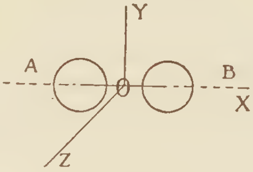


FIG. 76.

For **diatomic gases**, it is no longer permissible to neglect the energy due to rotation in comparison with the remainder of the kinetic energy. The molecules of these gases are considered to be dumb-bell like in shape. We must remember, however, that all directions cannot be regarded as equivalent axes, about which the rotations can occur. First, let us consider the axis AB, Fig. 76, which is formed by joining the centres of two atoms together. A rotation about this axis will only be equivalent to a rotation of each atom about its own centre—the energy will not be greater than when the two atoms rotate separately. Its contribution is therefore so small that it can be neglected. Suppose now that we consider a second axis SY which passes through the centre of gravity of the molecule and stands perpendicular to the first axis and lies in the plane XOY. The third axis SZ is perpendicular to both and therefore to the plane which contains them. It is possible for the energy of rotation about these two axes to be considerable. We have now to discover the magnitude of this energy. **Boltzmann** investigated the problem, and came to the simple conclusion that, in such a gas, a state of equilibrium is reached in which the kinetic energy due to rotation about one axis equals the kinetic energy due to that of a motion along one of the three possible directions. He expressed this result as follows: A gas molecule which can move in one of three directions has three degrees of freedom. With every degree of freedom there is associated an amount of energy equal to $RT/2$; when the rise in temperature is 1° C. the increase in energy for each degree of freedom is $R/2$, so that for three degrees of freedom it is $3R/2$. For diatomic molecules possible rotations give two more degrees of freedom—one for each axis about which an effective rotation takes place—so that the further increase in energy due to rotation is $2RT/2$. Hence, when the rise in temperature is 1° $C.$, this increase is $2R/2 = R$. It therefore follows that the total amount of energy necessary to heat one gram-molecule of a diatomic gas at constant volume through 1° $C.$ is

$$C_v = 3R/2 + 2R/2 = \frac{5R}{2} = 5 \text{ cals.}$$

Under similar conditions, but at constant pressure, a further amount of work equal to R must be done. We therefore have

$$C_p = 5R/2 + R = 7R/2 = 7 \text{ cals.},$$

and the ratio of the two specific heats for such a gas is

$$C_p : C_v = 7 : 5 = 1{\cdot}4 \quad . \quad . \quad . \quad . \quad (81)$$

It may also be possible for the two atoms of such a molecule to execute an oscillatory motion along the line AB joining their centres together. If this should occur we have another degree of freedom, which, if the same amount of energy is associated with it as with the others, causes the ratio of the specific heats to become still smaller.

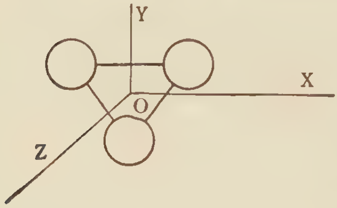


FIG. 77.

For a molecule consisting of three atoms, Fig. 77, we have to assume that the energy due to a rotation about a line joining the centres of two molecules is considerable, if we do not adopt the improbable suggestion that all three molecules lie in the same straight line. Or, if any three arbitrary axes are drawn to pass through the centre of gravity of the molecule, a rotation about any one of these directions as axis requires an amount of energy to be expended. It therefore follows that, in this instance, there are three degrees of freedom due to rotation, so that we have

$$C_v = 3R/2 + 3R/2 = 6R/2 = 6 \text{ cals.}$$

and

$$C_p = 6R/2 + 2R/2 = 4R = 8 \text{ cals.}$$

The ratio of the specific heats is

$$C_p : C_v = 8 : 6 = 1{\cdot}33 \quad . \quad . \quad . \quad . \quad (82)$$

which is still more nearly equal to unity.

For polyatomic gases it is difficult to say how many numbers of degrees of freedom they will possess. In addition to the energy due to their rectilinear motion, there is the energy due to rotations about all three axes, and perhaps some due to a possible vibration of some, or all of the atoms, about some zero positions. We can only say with certainty that (1) the difference between the two specific heats $C_p - C_v$ is always constant and equal to 2 calories, because all gases expand at equal rates. This is of course only true if our assumption is correct that the difference between these two values is due to the work done in overcoming the external pressure; (2) the ratio $C_p : C_v$ will more nearly approach unity, according to the measure in which the external work becomes negligible when compared with the internal energy of the gas. The values of this ratio will therefore decrease from a maximum value 1·66 for monatomic gases, to values which become smaller and tend to equal unity as the number of atoms per molecule is increased, but this limit will never be actually reached. The accompanying table contains the experimental values for gases which have been so far investigated. It begins with the rare gases and ends with some which are polyatomic.

It will be seen that the results contained therein are in complete harmony with our theoretical deductions.

Table 15

	Formula.	Gas.	Mol. Wt.	C_p.	C_v.	C_p/C_v.	$C_p - C_v$.
Monatomic	He	Helium	4	5·0	3·0	1·66	2·0
	A	Argon	39·9	5·07	3·07	1·65	2·0
Diatomic	H_2	Hydrogen	2	6·82	4·83	1·41	1·99
	O_2	Oxygen	32	7·0	5·0	1·40	2·0
	N_2	Nitrogen	28	6·98	4·98	1·40	2·0
	HCl	Hydrogen chloride	36·5	7·08	5·06	1·40	2·02
	CO	Carbon-monoxide	28·0	$7{\cdot}00_6$	5·00	1·40	2·00
		Air (without CO_2)	28·95	6·962	4·975	1·40	1·987
Triatomic	CO_2	Carbon dioxide	44	8·89	6·84	1·30	2·05
	N_2O	Nitrous oxide	44	9·66	7·54	1·28	2·12
	SO_2	Sulphur dioxide	64	9·86	7·64	1·29	2·22
Polyatomic	NH_3	Ammonia	17	8·84	6·75	1·31	2·09
	CH_4	Methane	16	8·48	6·47	1·31	2·01
	C_2H_6	Ethane	30	12·39	10·32	1·20	2·07

For the monatomic gases the agreement is practically perfect.

For diatomic gases the specific heats are greater by an amount required by the assumption that they possess two additional degrees of freedom. An arrangement of atoms in a gas molecule in which only four degrees of freedom exist cannot be imagined, *i.e.* it is impossible for a gas to exist in which $C_p = 6$ and $C_v = 4$. The absence of such a gas in the above table affords striking confirmation of our theory. In the remarkable agreements between theory and experiment which the above table manifests, even in its details, one can see not only a very strong confirmation of the kinetic theory as a whole, but also of the special views which form the basis from which it has been developed.

3. Atomic Heats

From the knowledge which has already been obtained, attempts have been made to obtain a deeper insight into the behaviour of solid and liquid bodies. For that purpose, instead of using the gram as the unit of mass in calculating specific heats, the gram-atom, or gram-molecule, has again been used. It is only necessary to multiply the specific heat by the atomic weight in order to obtain this quantity. **Dulong** and **Petit** first advocated this in 1819. They discovered the law which bears their name. The following table contains the atomic heats of a few solids. It shows that,

TABLE 16

Substance.	Atomic Weight.	Atomic Heat.
Aluminium	27	5·8
Lead	207	6·5
Iron	56	6·4
Gold	197	6·4
Nickel	58·7	6·4
Platinum	195	6·3
Mercury	200·6	6·4
Sulphur	32	5·7
Zinc	65	6·2

as a general rule, the atomic heat is approximately 6·4 cals. But the law is not a universal one. Apart from the fact that it is only approximately true for most elements, the discrepancies occurring in several instances are very great; for carbon, as diamond, the ratio is 1·35, as graphite 1·9; whilst boron gives 2·6, and silicon 4·8.

The atoms of solid substances are not free to move in all directions, but are constrained to move about a zero position about which point, in general, they can move in three directions. Hence, in addition to their kinetic energy they will possess a potential energy which, according to the remarks on p. 47, is equal to the kinetic energy. It is therefore to be assumed that there are six degrees of freedom for atoms in solids. If it is further conceded that each degree has an energy $RT/2$, then, in order to raise the temperature of a gram-atom of a solid, $6R/2 = 6$ cals. are necessary. In most instances this agrees very well with the observed values, but the considerable differences which had been found for the few elements cited above resulted in this subject being neglected for a long time. In 1875, however, **Weber** discovered that the low atomic heats of carbon, boron and silicon, increased with temperature so that their atomic heats approached those of other substances. He found that, for carbon, the atomic heat increased from 1·9 at ordinary temperatures, to 5·6 at 1000° *C*. This discovery led others to investigate the specific or atomic heats of normal substances at low temperatures.

The Atomic Heats at Low Temperatures. Experiment has revealed the very remarkable fact that the specific heat of all solid bodies becomes less as the temperature decreases. Foremost amongst those who have been engaged in these researches are **Nernst** and his pupils. Having designed special apparatus, they have investigated the specific heats of all possible substances at temperatures as near to the absolute zero as were feasible. They have been able to show that the specific heat of all solid matter becomes less and less as the absolute zero is approached, and appears to become zero in the immediate neighbourhood of that point. **Debye** assumes that at these very low temperatures, little by little, the vibrations of the individual atoms with respect to one another cease and that only large groups of atoms, and finally pieces of the substance itself, oscillate together. From this it follows that the number of particles which move independently of one another becomes less. Since, however, the same amount of energy is necessary to raise the temperature of a molecule however large the mass of vibrating particles may be, it follows that the specific heat must decrease rapidly.

Nernst and many of his pupils have collaborated in investigating the specific heat of gases at low temperatures. Their results may be summarized as below.

Monatomic gases, which have been discussed above, possess three degrees of freedom from the lowest to the highest temperatures; the specific heats are $C_p = 5$ and $C_v = 3$ cals. It is only when the temperature has been raised to such an extent that the particles begin to radiate energy, that the specific heat begins to increase by an amount which depends upon the magnitude of the radiation. This, however, cannot be considered to be a real change in the specific heat of the body, for the raising of the temperature has caused another phenomenon to appear, viz. the gas has begun to emit radiation. Here, however, we can disregard this radiation.

For **diatomic gases**, with their five degrees of freedom, we have, according to theory, $C_v = 5$ and $C_p = 7$ cals. From Table 16 we see that this is true, for most gases in this category; but the extension of observations to high and low temperatures has revealed the fact that these values are only valid over a few hundred degrees range in temperature. At high temperatures the specific heat of all diatomic gases increases, but one must remember that at these temperatures the gases begin to undergo a further change, for the chemical bonds begin to break down, *i.e.* the molecules dissociate into atoms. This breaking of the bond necessitates the expenditure of energy, the amount of which is not negligible. But long before the dissociation temperature has been reached the specific heat has begun to increase, and with gases such as chlorine, Cl_2, which dissociate at lower temperatures, it is found that they have a higher specific heat at temperatures where other diatomic gases still behave quite normally. We have to imagine that the ultimate breaking down of a molecule is preceded by a vibration of the atoms around their common centre of gravity—this will correspond to an additional degree of freedom.

At the lowest temperatures it has also been found that the specific heats are no longer equal to their theoretical values. **Eucken**, who investigated the specific heat of hydrogen down to $-240°$ *C.*, proved that the specific heat decreased continuously as temperatures below 0° *C.* were reached, and at 35° absolute temperature he found $C_v = 2{\cdot}98$ **a value which is equal to that for monatomic gases.** It therefore follows that at this temperature two degrees of freedom have been lost. These are obviously the two degrees due to rotation so that the only degrees of freedom which remain are those due to translation; the hydrogen molecule is therefore behaving like a monatomic gas with three degrees of freedom.

For gases, the molecules of which are composed of several atoms, phenomena of exactly the same type occur. But at high temperatures, when these constituent atoms begin to vibrate, there will be a large increase in the number of degrees of freedom, so that the corresponding increase in specific heat is more rapid than in the case of diatomic molecules. When these gases are investigated at temperatures near to that of the absolute zero their molecules all cease to rotate, so that these gases behave like monatomic or diatomic gases, and only possess three degrees of freedom. For such molecules $C_v = 3$, and this value is approached when polyatomic gases are examined at these low temperatures.

4. THE QUANTUM THEORY AND ATOMIC HEATS

One of the greatest difficulties which confronted physicists at the beginning of this century was to offer an explanation of the fact that the specific heat of solid bodies decreases as the temperature is diminished. The

Quantum Theory which had been enunciated by **Planck** (1900) was applied by **Einstein**, in 1907, to solve this question. Later on we shall find occasion to discuss this theory more fully. Here we shall only state those few ideas which are necessary in order to appreciate the problem under discussion. The fundamental idea of Planck is that energy has a kind of atomic structure—*i.e.* energy is really made up of tiny 'packets', or quanta of energy, so that wherever energy is absorbed, or emitted, at least one quantum must be involved. Planck had only used this idea in connexion with the radiation of heat. Einstein extended the idea so as to incorporate the energy due to the vibrations of the particles in solid bodies, *i.e.* to heat itself. Now according to Planck's hypothesis a quantum of energy is not invariable, but depends upon the frequency of the moving particle so that the smallest amount of energy which can ever be involved is $e = h\nu$, where ν is the frequency and h a universal constant of nature. The frequency, in the case of solids, is determined by the forces which hold the molecules together, for that is also the force which tends to make the molecules return to their original position. In general this force is not appreciably affected by changes in temperature, *i.e.* its magnitude will not be very different at low temperatures from what it is at high temperatures. Hence the quantum of energy which can be taken up by a solid body at low temperatures is not smaller than at high temperatures. Gaseous substances do not behave like this, for the lower the temperature the less is the energy which can be associated with their molecules. We therefore have to imagine that a solid body, the molecules of which cannot move freely from their positions, is heated by the impacts of the molecules of the gas by which it is surrounded. The lower the temperature the more frequently will it happen that the energy given up by a colliding particle is less than the quantum of energy which the solid body can receive. Hence, at low temperatures, there will always be present in a solid body more and more particles to which energy is not being supplied by collision. In consequence of this they will not vibrate, a fact which is equivalent to saying that the number of degrees of freedom diminishes when the temperature is lowered. At the lowest possible temperature only a very few particles will receive a sufficiently large quantum of energy. These views of Einstein were immediately supported by the fact that it is in diamond, which is known to be one of the hardest substances, that the largest discrepancies are found, and even at room temperatures its specific heat is very low. The fact that it is so hard means that its atoms are bound together by very large forces so that their frequency must also be very great and the quantum of energy associated with them must also be one of the largest. This explains the fact that even at ordinary temperatures the quanta of energy are not sufficiently big to excite the atoms, so that the solid body will no longer receive energy although all its atoms are not yet moving with the higher velocity. It therefore receives less energy than a body in which the frequency of vibration is less. These views of Einstein have been embodied in a formula which agrees well with the variations in the specific heat, as found by **Nernst** and his pupils.

Conclusion. If we now revise the contents of this chapter it will be found that there are three stages which confront us with varying degrees of certainty. The first section is comprised by facts which have been observed and are irrefutable. In the front rank of these are the numerical values of the specific heats of different substances which, apart from incidental experimental errors, are absolute facts. In particular, there are the observed changes in the specific heats as the temperature is increased or diminished. The kinetic theory of gases endeavours to explain these facts. Since it must also be counted amongst observed facts, the additional idea of degrees of freedom becomes more firmly established when it is found that the facts are supported by theory. The old classical theory which had been brought almost to perfection by **Boltzmann** in Vienna was not able to explain all these phenomena. It was not that the phenomena contradicted the fundamental assumptions of the kinetic theory, but that a few phenomena could not be explained in terms of that theory; a few additional conceptions, which the old classical theory could not supply, were needed. These conceptions were given by the quantum theory, which had been developed by Planck, and which was used by Einstein to explain the phenomenon which we have discussed above. Later on we shall learn more about the quantum theory and radiation. At present we have to remain content with the fact that this theory has made many valuable contributions to the subject of specific heat, but these investigations cannot yet be regarded as having reached finality.

CHAPTER VII

HEAT AND THE VARIOUS STATES OF AGGREGATION

1. Preliminary Remarks on the Change of State

We have seen that heat is due to the more or less vigorous movements of the smallest particles which constitute bodies, and we also know that even in solid bodies the molecules are not held together by means of infinite forces, since all solid bodies can be divided into smaller portions. The fact that the shape of a body becomes endangered when the molecular movements increase, and is finally lost when they have become sufficiently vigorous, is one which ought no longer to make us wonder.

In all the states of aggregation there are attracting forces between neighbouring molecules and atoms. These are opposed by forces which are the results of the thermal agitations of the molecules. In solid substances there is a continuous exchange between the two, just as in a vibrating body there is a continuous exchange between its inertia and the elastic forces. As the distance from the zero position increases, however, the force of attraction between the molecules which is also the force tending to bring them back to their zero position must diminish in intensity. When the amplitude of vibration has reached a certain value, in other words, when the temperature of the body is sufficiently high, the above forces are not capable of arresting the motion of a particle and causing it to return to its zero position. The particle will therefore be permanently removed from its position of rest. Since the temperatures inside a body are nowhere very different from one another, this condition will be reached by all the molecules at the same time, *i.e.* there is a complete collapse of the shape of the body which has hitherto existed—in other words, the body **melts.**

When the liquid state has been reached the molecules are no longer constrained by forces tending to hold them in some zero position, but nevertheless they still remain together. The collisions are not sufficiently powerful to separate the molecules from one another completely. In the interior of the liquid the forces between neighbouring molecules are in all directions and are equal in magnitude so that their resultant is zero. Also, when two molecules collide, the kinetic energy is so great that it cannot all be converted into potential energy by the attractive forces, when the molecules are receding from each other after collision. But at the surface of a liquid the molecules only experience a pull towards one side, *i.e.* the interior of the liquid ; there the forces of attraction are greater than those which tend to split up the molecule. Hence liquids do not have a fixed shape but only a definite volume, and an easily distinguishable surface which separates them from other bodies. Under the action of these forces small masses of water assume a spherical shape. If the temperature rises the molecular velocities become greater, and finally the forces at the surface are not sufficient to hold the particles together. These particles move further and further from their late neighbours, their motions no longer being impeded by the forces of attraction; occasionally they will collide with other molecules and thereby alter the direction in which they are moving; the molecules will have passed over into the gaseous state.

In dense gases such collisions will occur very frequently; but the essential characteristic of a gas is not the collisions which occur, for these only take place less and less frequently as the density diminishes, but the large velocity with which the molecules move. These velocities are so large that the forces of attraction, even when the molecules happen to get very close to one another, pale into insignificance when compared with them.

The molecules of most gases are not indivisible entities, but consist of several atoms, like H_2, O_2, HCl, CO_2, CH_4. These atoms stick together very much like the atoms of a solid body. When the temperature is very high, the phenomena which are associated with the specific heat of a body, and which we have already discussed, cause us to assume that the atoms of the gaseous molecules begin to oscillate about their positions of equilibrium. In a collision where preferably one atom of the molecule is hit, the latter will either be made to approach or separate from the remaining atoms of the molecule; in so doing it will of course be removed from its position of equilibrium and will follow the impulse which tends to restore it to this position, thereby commencing to oscillate about this zero position. Since the forces which hold the atoms of a molecule together are only finite (a fact which is revealed by many chemical processes), we have to expect that when the temperature is sufficiently high the molecule will begin to decompose further. This is what we can actually observe. At high temperatures all gases are decomposed into single atoms, *i.e.* they are dissociated. But since, according to the kinetic theory, the velocities of all the molecules are not strictly speaking the same, it follows that dissociation will only take place amongst those molecules, the velocity of which is sufficiently high. As the temperature increases the proportion of molecules having this velocity will increase, so that the particular temperature at which dissociation occurs cannot be stated definitely.

Have we now reached a limit beyond which matter cannot be decomposed any further? We certainly should if the atom is either a simple entity, or something whose parts are so strongly held together that there are no means at our disposal whereby they can be separated. We certainly have no means whereby the temperature can be indefinitely increased. But if the only condition for the decomposition of matter is that the atoms should be hit at one place by a great force, then Nature herself provides these forces. In dealing with the phenomena of radioactivity we shall learn about the small atoms of helium (alpha particles), which are hurled forth from radioactive bodies with a velocity so large that it equals $\frac{1}{15}$ that of the velocity of light. Recently (1919) the experiments of **Sir Ernest Rutherford** have taught us that it is possible to use these small atoms as projectiles to decompose various atoms. At first the experiments were made on nitrogen. If we wish to include these phenomena in our considerations we can easily calculate the temperature to which a quantity of nitrogen would have to be raised in order for the momentum of its own atoms to break up others into their constituents. The velocity of an alpha particle is different according to the parent from which it is derived. Let us assume a mean value of 2×10^7 m/sec.

The velocity of the same particle as an atom of helium at 0° *C.*, *i.e.*

$273° K.$, is about 1200 m/sec. Hence by using equation (72), which states that

$$\frac{T}{T_0} = \frac{v^2}{v_0^2},$$

we have, in our particular instance,

$$T = 273 \times \frac{4 \times 10^{14}}{144 \times 10^4} = 7{\cdot}6 \times 10^{10}.$$

Since the momentum of all molecules is the same at this temperature we should certainly expect that, at a temperature of $7{\cdot}6 \times 10^{10}$, the atoms of nitrogen would become broken down into smaller particles : this would not only happen with a few of the molecules, as in Rutherford's experiments, but the whole mass, in consequence of the momentum of the molecules at that temperature, would be disintegrated. In other words, nitrogen and the various other elements could no longer exist at this temperature. The fundamental fact upon which this theory is based is that if we permit particles having this momentum to be hurled at atoms, then these latter are actually found to be destroyed. Later on we shall return to discuss this question again.

In conclusion, if we once more survey the whole series of ideas which have been given above, we behold an unbroken chain of facts, from the stone-breaker at the roadside who breaks the stones with his hammer, to the disintegration of an atom by means of an alpha particle.

2. FUSION

Now that we have to discuss the various processes individually we shall first record the actual facts and then show how these may be accounted for in terms of the kinetic theory. It has been found possible to convert all solids into liquids and gases. On the other hand, all substances which are gaseous at ordinary temperatures have been liquefied and solidified. For purposes of illustration the following table records the transition points for some of the more common substances. The melting point is the temperature at which solids become liquid. For the sake of completeness the boiling points are also quoted ; these represent the temperatures at which liquids are converted into vapours. The boiling points are influenced by the pressure which prevails when the determination is carried out so that the figures quoted assume the pressure to be that of a standard atmosphere.

TABLE 17

Substance.	M.P.	B.P.	Substance.	M.P.	B.P.
Helium	− 272·3	− 269	Tin	232	2270
Hydrogen	− 259	− 252·7	Lead	327·4	1525
Oxygen	− 250	− 183	Aluminium . . .	658	1800
Nitrogen	− 210	− 196	Sodium chloride . .	801	1440
Argon	− 189	− 186	Copper	1083	2300
Chlorine	− 100	− 34·5	Iron	1530	2450
Mercury	− 39	+ 356·7	Platinum	1764	3800
Water	± 0·0	100	Fused Quartz . .	1700	2200
Acetic Acid . . .	+ 16·6	118·5	Osmium	2700	—
Sulphur (rhombic) .	113	444·5	Tungsten	3400	4800
Sulphur (monoclinic)	119		Carbon	3600	3950

The melting points have a range of temperature from − 272·3° C. (only 0·9° above the absolute zero) for helium, to 3400° C. for tungsten.

It has not yet been possible to liquefy carbon. In 1913, however, some experiments were conducted in which carbon was heated at a very high temperature and under a very high pressure. When the pressure was released the carbon was found to be in the form of drops. From this fact it was concluded that the carbon had been liquid when subject to the high pressure. Recently the same phenomenon has been observed with carbon at about 3600° *C.* and under ordinary pressure.

The Heats of Fusion. A mixture of snow and water which has been well stirred has a steady temperature—it has been defined as 0° *C.* Even when the mixture is warmed by a flame the temperature does not rise at once, but the snow melts more rapidly. It is only when the snow has all melted that the temperature begins to rise. But if the mixture is not heated by a flame and neither receives nor imparts heat to its surroundings, then the snow and water remain together unchanged. In order to melt the snow (or any other substance) heat is necessary. This quantity of heat can be found by investigating how much water at 100° *C.* must be added to 100 grm. of snow at 0° *C.* in order for the mixture to be water at 0° *C.* It is found that 80 grm. of water at 100° *C.* are required. If the heat required to melt one gram of snow is L cals., then for 100 grm. it is 100 L. Now according to the experiment this is equal to the heat given out by 80 grm. of water in cooling from 100° to 0° *C.*, so that $80 \times 100 = 100\ L$ or $L = 80$ cals. This is a very considerable amount of heat. In order to melt one gram of ice the same amount of heat is necessary as would raise the temperature of one gram of water from 0° to 80° *C.* This heat ceases to exist as such after the ice has melted. This heat is therefore called the latent heat, but this is an expression which it is better to avoid.

The heat necessary to melt a substance is only latent in much the same way as the kinetic energy of a stone, which has been projected upwards, has become latent at the instant when the stone is at its maximum height. For the kinetic energy has then vanished, or if one wishes to use the expression, it has become latent. But in place of it we have a corresponding amount of potential energy. Likewise in the case of liquids and solids, for the former are to be distinguished from the latter by the fact that their molecules possess potential energy with respect to those in the solid state, and just as the kinetic energy again becomes manifest when the stone falls, so the heat which has disappeared is recovered when the substance passes into the solid state again. The following table gives a list of substances and their heats of fusion.

TABLE 18

Heats of Fusion

Element.	Heat of fusion in grm.-cals. per grm.	Compound.	Heat of fusion in grm.-cals. per grm.
Aluminium	94	Ice	79·7
Lead	5·5	Ether	27
Iron	49	Benzene	30·4
Gold	15·9	Chloroform	47
Platinum	27	Glycerine	42·5
Silver	10	Acetic Acid	43·7
Zinc	23	Ammonia	108·1
Mercury	2·8	Sulphuric Acid	8·8

The above table shows that the heat of fusion of ice is greater than that of most other substances, and this fact must show itself in a variety of ways in Nature, when ice melts or water freezes.

The Change in Density when Substances are Melted. When a substance changes from a solid to a liquid there is in general a small change in density: most bodies expand on melting, but a few of them contract. Water belongs to this second class of substance; ice has a density which is less than that of water, and so floats on the surface of water. The difference is quite considerable for it amounts to about 9 per cent. It is on account of this that icebergs float with about one-tenth of their mass out of the water. Besides water, bismuth, antimony, and pig-iron expand on becoming solid. It is reasonable to suppose that, at the moment* when a body freezes, the atoms take up a definite arrangement and that this accounts for any expansion or contraction which may take place, but until recently our knowledge of this arrangement was very limited. Nowadays the arrangement of the atoms in crystals can be determined by the use of X-rays, and this knowledge has provided us with the necessary clue in order to solve the problem.

The fact that there is an alteration in volume when a change of state occurs necessitates that the melting and boiling point of a substance must depend upon the pressure to which they are subjected. If, for example, a substance expands on melting, as most substances do, then a greater amount of work will have to be expended at the moment of melting if the pressure is increased. A temperature which is sufficient for fusion to occur under atmospheric pressure is not sufficiently high when the pressure is increased. On the contrary, when bodies contract on melting, the contraction implies an increase in energy equal to $P\Delta V$ and this energy increases with the pressure. It is therefore not necessary to supply heat by raising the temperature, so that when a substance like ice melts under increased pressure, it does so at a lower temperature. (In the case of a glacier the ice melts at those points where it is actually in contact with the earth and pressing against it, so that the glacier moves along more rapidly.) By applying the laws of thermodynamics to this problem it can be shown that when the pressure is increased by one atmosphere the freezing point of water is lowered by $0{\cdot}0072°$ C. Hence, in order to lower the melting point $1°$ C. the pressure must be increased by 133 atmospheres.

3. EVAPORATION

It is a well-known fact that a mass of water slowly disappears when it is left exposed to the open air. The pools of water, found in the streets or lanes after a storm, dry up more quickly in a wind than when it is calm; they also disappear more rapidly if they have been formed in an open field, rather than in a wood. By means of a sensitive thermometer it can very easily be shown that the temperature of the water falls when water thus disappears. If the bulb of a thermometer is covered with damp cotton wool the temperature indicated by the thermometer diminishes, especially when the thermometer is in rapid motion. If the cotton wool is moistened with ether instead of water then, even in summer, the temperature may fall below $0°$ C. and frost appear on the wool. The drying of wet clothes by hanging them in an open space is another example of the disappearance of water by a process which is called evaporation.

---

* It is only pure substances, pure metals, and eutectics, which freeze at definite temperatures. [*Trans.*]

All liquids evaporate at all temperatures, but at very different rates; the higher the temperature the more quickly does the evaporation take place. With different liquids the evaporation is greater as their temperature approaches that at which they boil, *i.e.* at room temperatures, liquids which have a low boiling point evaporate more readily than others.

To-day we are quite accustomed to regard this disappearance of a liquid as being due to its passage into a gaseous state. Why is it that this which is so obvious to us should have presented such difficulties to the ancient philosophers? On several other occasions we could have raised similar questions, but here the question becomes more acute because, from the fact that water vapour is formed from liquid water, the ancients believed that they had found a support for their doctrine concerning the transmutation of matter. Why were these ancient people able to form such a doctrine and why is it that to-day we hold such different views? It is necessary for us to stop a moment and consider the facts and theories with which these ancient philosophers were familiar, for it is very difficult for us to renounce completely the things we have learnt as children and which familiarity has fixed in our minds. The chemical similarity of the liquid which had disappeared and the vapour which had been formed was not known to these people. They only saw in liquids and vapours their physical properties as we should call them to-day. And surely these were sufficiently different: the heaviness of water and the lightness of its vapour; the surface of a liquid which enabled it to be seen and the absence of such a surface in the case of all the gases which were known at that time; their mobilities too were very different, and we must remember the important rôle of motion in these far-off days. To them it was obvious that a fundamental change took place when a substance passed from one state of aggregation to another. Moreover, they had only observed this transition in connexion with a few substances. The view that there were solid and gaseous bodies, in the sense that these states were inherent properties of the bodies, has only really disappeared during the last fifty years, when it has been shown that all the three states of matter can pass from one to the other, *i.e.* all gases have been liquefied and all solid smelted. But so long as there was an essential difference between solid and liquid bodies, the transition from one state to another was a fundamental natural change.

To-day we look upon these phenomena from quite another point of view. We know that the nature of a substance is not to be discerned by a superficial examination of it, so that we do not form conclusions as to its nature from a few observations on phenomena which present themselves to us, for we have learnt, only too often, that such conclusions do not always stand the test of time. On the other hand, we know that a change from one state to another is a very common occurrence; it happens thousands of times every hour of the day, at all places, and with very different substances. The formation of rain, and the subsequent drying of the roads after rain, are examples which are continually being repeated. The fact that the properties of the water and its vapour are similar refutes the idea of there being any fundamental change in their nature, and above all, the investigation of the laws which govern their formation, and the conditions under which the transition from one state to another occurs, have shown us quite clearly that we are only dealing with a different form of the same substance.

The Explanation of these Phenomena. Since we are no longer able to doubt the validity of the kinetic theory we must try to interpret the above phenomena according to its postulates. This can be done very satisfactorily.

The velocity of the molecules in gases and liquids which we use in our calculations is only a mean value. The actual velocities are in some instances above and in others below this value. Using the theory of probability, **Maxwell** discovered the rule which tells us how these actual velocities are distributed around their mean value. The greater the divergence from the mean value the less is the number of molecules which move with this velocity. There are always present, however, a few molecules which will have a velocity which is much greater than the mean velocity. In other words, their kinetic energy is so large that they will have passed beyond the region of capillary attraction before their velocity has become zero. The number of molecules having such a velocity will become greater as the temperature of the liquid rises. Since it is only these rapidly moving molecules which leave the liquid, it follows that the mean velocity of the remaining molecules must have diminished, *i.e.* the liquid is cooled in virtue of this process, and the heat of vaporization is a consequence of this process, which can be very easily understood. When the gaseous molecules get into the open air above the liquid they will be removed from this position very quickly if the air is moving rapidly. If, on the other hand, the liquid only half fills a hermetically sealed vessel, the molecules which have escaped from the liquid must remain in the space above it. In consequence of the uneven distribution of velocities amongst the molecules of this vapour it follows that some of these molecules will return to the liquid; the number which do return will increase as the density of the vapour becomes greater. The process of evaporation into a closed space is therefore accompanied by another phenomenon tending to retard the evaporation; for since there is present in the liquid, even at the same or a somewhat lower temperature, only a definite number of molecules with sufficient velocity, and since the number of molecules which return to the liquid becomes greater as time goes on, it follows that a state of equilibrium will be reached when the number of molecules leaving the liquid per second is equal to the number which returns per second. So long as the temperature does not change, the mean value of this number will remain constant. In other words, the liquid will be in contact with its own vapour, which, at a definite temperature, has a constant density. If we endeavour to alter this density, let us say, by diminishing its volume, which can be done by increasing the volume of liquid present, or else by forcing the piston in Fig. 78 downwards, then the number of particles returned to the liquid would be increased, whereas the number of particles coming from the liquid would remain as before. Ultimately the density of the vapour would be at its original value. This is a fact which is observed whenever a liquid is in contact with its vapour in a closed space.

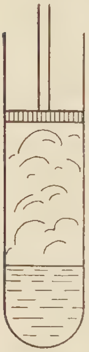


Fig. 78.

But if the temperature of the liquid is increased then the molecules leaving the liquid would increase in number so that the density of the vapour would also become greater. Finally, equilibrium would be attained when the number of molecules entering the space above the liquid is equal to those which leave that space. This state of equilibrium will only be reached when the density of the vapour is greater than it was at the lower temperature. It is therefore necessary to ascribe to every liquid a definite vapour pressure, which increases very rapidly with temperature.

From this hypothesis it does not follow that the state of equilibrium is characterized by the fact that the process of evaporation has come to an end, and that all is finished ; all that it does imply is that the two processes still go on with the same mean velocities so that the resultant effect is the same. Such a state of equilibrium is said to be a dynamic one, in contrast to a static one, in which the resultant force is actually zero so that nothing further happens. In our problem the forces are not directed against one another, but it is only the resultant effect which remains constant ; the one process is tending to increase the density of the gas whilst the other tends to diminish it.

This vapour pressure can most easily be measured in the following way. The liquid whose vapour pressure is required is placed in sufficient but otherwise quite arbitrary quantity above the surface of mercury in a barometer. The barometer tubes shown in Fig. 79 are as follows : (*a*) is free from air in the space above the mercury, whilst in (*b*), (*c*) and (*d*) water, alcohol and ether have been placed respectively. It is now no longer the mercury column alone which is in equilibrium with the pressure of the atmosphere, but the column of mercury together with the pressure of the superincumbent vapour. Consequently the level of the mercury is lower than in (*a*). If the pressure of the atmosphere is H, whilst B is the length of the mercury column in b, then the vapour pressure P of the water at the temperature of the experiment can be calculated from the equation $P + B = H$, *i.e.* $P = H - B$.

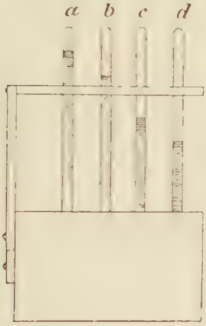


FIG. 79.

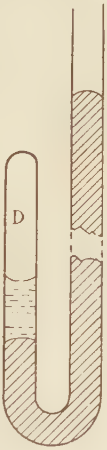


FIG. 80.

Proceeding in this way the vapour pressure of water up to 100° C. can be determined. At this temperature its vapour pressure becomes equal to that of a standard atmosphere, so that $B = 0$. In order to measure the vapour pressure at temperatures about 100° C. some other form of apparatus must be used—see Fig. 80. When the temperature increases the mercury will rise in the right-hand limb of the tube. The vapour in the space D is then under a pressure equal to that of the atmosphere plus that exerted by the mercury column, so that $P = B + H$.

The following table contains the vapour pressure of water at different temperatures. Until a pressure of one atmosphere is reached the pressures are given in mm. of mercury. The higher pressures are given in terms of the atmosphere as unit (1 atmos. = 760 mm. of mercury).

Table 19

The Saturation Vapour Pressure of Water at different Temperatures

T ° *C.*	Pressure (mm.).	*T* ° *C.*	Pressure (mm.).	*T* ° *C.*	Pressure (mm.).	*T* ° *C.*	Pressure (mm.).	*T* ° *C.*	Pressure (atmos.).
− 60	0·007	0	4·6	60	149·4	90	525·0	100	1
− 50	0·029	10	9·2	65	187·5	92	567·0	120	2
− 40	0·093	20	17·5	70	233·7	94	610·9	134	3
− 30	0·280	30	31·8	75	289·1	96	657·6	153	5
− 20	0·77	40	55·3	80	355·1	98	707·3	180	10
− 10	1·95	50	92·5	85	433·6	99	733·2	200	15·3

Sublimation. An examination of the above table shows that even at temperatures below 0° *C.* vapour pressures have been recorded. This means that even ice is surrounded by water vapour. If this vapour is continuously being removed from the region above the ice, this latter will slowly pass from the solid into the gaseous state, without ever becoming liquid. This process is called **sublimation.** In this way it is possible for a thin layer of snow to disappear without the temperature ever becoming greater than 0° *C.* if the vapour above the snow is continually removed.

Since evaporation only depends upon the temperature of the liquid, and the return process of becoming liquid again only depends upon the number of molecules of vapour present in the vessel above the liquid, it follows that it is immaterial as to whether the space above the liquid contains any other gas or not; evaporation will continue until the pressure has reached the appropriate value for the liquid at that temperature. The space is then said to be saturated with the vapour. Thus whenever the air of our atmosphere is in contact with a large surface of water (clouds and fog must be considered in this category too) the air can contain a quantity of water such that its pressure is equal to that which water at that temperature would exert if it were only in contact with its own vapour, *e.g.* at 20° *C.* it would amount to 17·5 mm. of mercury. Now it is not exactly wrong to say that from then onwards the air is saturated with water vapour, but it would be a misinterpretation of the conditions if one believed that by saying ‘ the air is saturated with water vapour ’ the condition of the air had been definitely fixed. It is quite immaterial whether the space is filled with air or not, it is only a question of the temperature, and if for any reason this should fall, then the air, or more strictly the space above the water, would contain more water than it could hold at this lower temperature ; a portion would therefore condense until the pressure was such that it was consistent with the new conditions of temperature ; *e.g.* at 20° *C.* the pressure of the water vapour may be as much as 17·5 mm., but at 10° *C.* its greatest value is only about one half of this, viz. 9·2 mm. Hence, when air saturated with water vapour at 20° *C.* is cooled to 10° *C.*, half its water must be precipitated. This water is generally to be seen in the form of clouds (these consist of such tiny drops of water that they are able to remain suspended in the air, just like the motes we see in a sunbeam). When these drops become too large they fall as rain to the earth. In winter, when the air is cooler, it contains less water vapour than in summer.

4. BOILING

The surface layers of a quantity of water contained in a beaker are always under a pressure equal to that due to the prevailing atmospheric conditions. At points which are inside the liquid the pressure is increased owing to the presence of the water column which stands above it. When the water in such a vessel is heated from below, its lower layers always have a temperature which is greater than that elsewhere. In consequence of this the density of the water here becomes less, so that this water rises, and is replaced by cooler water from above ; this causes the water to be mixed thoroughly, so that no great temperature differences can exist anywhere ; but the temperature is always greater in the lower layers of the water, and it is there that the molecules first attain the velocity which enables them to escape from the liquid.

Since it is always the particles with the highest velocity that leave the water, it follows that the temperature of the latter must fall, so that when evaporation is occurring more heat energy must be supplied to the liquid if it is to maintain its temperature. This energy is called the **heat of vaporization**. When the vapour becomes liquid again this energy is given up and appears as heat. In the following table the heats of vaporization for a few liquids are recorded.

TABLE 20

Heats of Vaporization at Atmospheric Pressure

(gram-cals. per 1 grm. of substance.)

Liquid.	Heat of Vaporization.
Ether	90
Alcohol	202
Benzine	94
Bromine	43
Carbon bi-sulphide	85
Turpentine	70
Mercury	68
Water	539

When a large mass of water is at such a temperature that vapour at atmospheric pressure can be formed, small visible bubbles of vapour appear inside the liquid. Strictly speaking these bubbles are formed under a pressure equal to atmospheric together with the small pressure arising from the column of water above them. This pressure will cause them to be compressed until the pressure inside them is equal to this external pressure. If we neglect this small excess pressure we can say that the bubbles which rise have a pressure equal to atmospheric. Initially, the cooler regions of the water will cause the temperature of these bubbles to fall so that their pressure is no longer sufficient for them to be in equilibrium with that of the atmosphere. Therefore, having commenced to rise, they will disappear, for they will have condensed. When the temperature of these upper layers has become high enough for vapour at atmospheric pressure to be formed from it, the bubbles will pass through it without condensation into the open space above. The formation of bubbles in the interior of the water now becomes continuous and the water is said to **boil**. As soon as the temperature at some point in the liquid, owing to some local heating effect, is a little higher than elsewhere, bubbles will be formed there. These bubbles carry some of the excess heat with them so that the temperature falls until the vapour pressure is again atmospheric. If the external pressure should vary, let us say increase, then the temperature is no longer that at which vapour at this pressure can be produced. Bubbles will cease to appear, until sufficient heat has been supplied to the liquid to raise its temperature, so that its vapour pressure again becomes equal to the external pressure. When this has been done bubbles will again appear, but their temperature will of course be greater than before. If the pressure falls, then the water will boil at a lower temperature. If a sensitive thermometer is used to measure the temperature of the vapour over boiling water, then the pressure under which the boiling takes place can be calculated; from Table 19, on p. 170, we see that when water boils at 99° *C.* instead of at 100° *C.*, the external pressure is 733·2 mm.; or, in the neighbourhood of normal atmospheric pressure 1 mm. change in pressure causes a corresponding change of 0·037° in the boiling point.

At the same time we see that by heating water in a closed vessel enormously high pressures can be attained; ultimately they will become so large that any vessel will be broken.

5. CRITICAL TEMPERATURE AND PRESSURE

When water in an enclosed chamber is heated the water expands continuously so that its density becomes less and less; on the other hand the density of the vapour which is formed will increase continuously. Finally, if the temperature is raised sufficiently, the density of the water will equal that of the vapour above it. Hence, at this critical temperature, no difference exists between the liquid and its vapour; the liquid has no distinct boundary. Above this temperature water can only exist as a gas. We therefore see that, however much the pressure may be increased, no liquefaction can take place. This fact was of very great importance to those who had been trying to liquefy all gases. For a long time previously, many unsuccessful attempts had been made to liquefy air, hydrogen, nitrogen and oxygen, by increasing the pressure. When the significance of the critical temperature had been appreciated other pieces of apparatus were constructed. In these the gas was first subjected to the lowest possible temperature, the pressure being afterwards increased. In this way all gases have been liquefied, so that to-day we no longer know of any gas which cannot be converted into a liquid. The attainment of this object was realized with certain gases only by conducting the experiment at a temperature in the neighbourhood of the absolute zero. The tension of a vapour in contact with its liquid at the critical temperature is called the critical pressure. For carbon di-oxide the critical temperature is at + 31° *C.*; it is easily determined by placing an hermetically sealed glass tube containing liquid carbon di-oxide in a beaker of water [see Fig. 81]. By raising the temperature gradually all the phenomena which we have mentioned above can be experimentally confirmed. The following table gives the values of the critical temperatures and pressures for a few other substances.

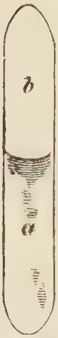


FIG. 81.

TABLE 21

Critical Temperatures and Pressures

Substance.	Critical temp. (° *C.*).	Critical press. (atmos.).
Helium	− 268	2
Neon	− 228	20
Argon	− 122	36
Krypton	− 63	41
Xenon	+ 15	43
Radium-emanation	+ 104·5	—
Hydrogen	− 240	10
Nitrogen	− 147	25
Oxygen	− 119	38
Methane CH_4	− 82	35
Ethane C_2H_6	+ 32	37
Propane C_3H_8	+ 97	34
Butane C_4H_{10}	+ 150	28

Attention is called to the values for the inert gases which show a regular increase in the critical temperature from helium to radium-emanation. A similar regularity occurs in the case of the hydrocarbons from methane to butane.

CHAPTER VIII

THE LAWS OF THERMODYNAMICS

1. The Conservation of Energy and the First Law of Thermodynamics

THE law of the conservation of energy has already been mentioned in connexion with mechanical processes, *i.e.* the energy was due to motion, gravitation, elasticity, surface tension, etc. It was then found that if one of these decreased by a definite amount during some process, then it must have been accompanied by an increase in energy of one of the other types; the increase is always equal to the decrease.

Expressed in the above form, however, this generalization needs further comment. It may well be asked, 'What right have we to say two different kinds of energy have been compared with one another and found equal?' The comparison is generally only possible when the second type of energy can be converted into the first type again. Then we compare the original energy with that which, after a second change, has been converted into energy of the original type. Let us assume, for example, that we have a pendulum, or some other body, vibrating. Initially we will suppose that the body has been displaced from its zero position to a position C (Fig. 82). It then has potential energy. If the body is released it falls, the potential energy gradually diminishing until, when the body is at B, the only energy present is kinetic. We say that this kinetic energy is equal to the potential energy which has been lost. On what basis is this assertion made? How can potential energy be measured in terms of kinetic? The basis of our argument lies in the fact that kinetic energy can be converted into potential energy again. In the present problem, the agent which causes this conversion to occur is the pendulum itself, for, after the pendulum has reached its lowest position, it begins to rise again, and in so doing it loses its kinetic energy which reappears as potential energy. Eventually the pendulum regains its initial height by arriving at E. This is the important point in the argument, for we no longer compare different things with one another but things which are of the same type. In the present example we compare the potential energy at the beginning with the potential energy at the end of a half-period; actually we compare the height of C above B with the height of E above the same zero position. From the fact that these two heights are equal, we conclude that the kinetic energy is able to make good the loss in potential energy, and say that the loss in potential energy in falling from C to B is equal to the kinetic energy at B.

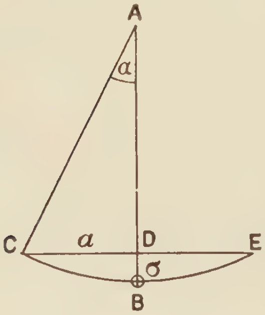


Fig. 82.

It has been shown that heat phenomena are due to molecular agitations; if therefore heat is due to a mechanical process one naturally assumes that the law of the conservation of energy might, or rather must, apply to all heat phenomena.

In spite of this the energy law must be proved—

1. The discovery of the laws of energy was made at a time when people were not yet convinced that heat was due to the kinetic energy arising from molecular movements; at all events the discoverer **J. Robert Mayer** (1814–1878) scarcely regarded heat as anything more than a condition of a body which produces the sensation of hot or cold when the body is touched, and it was necessary to spend a great deal of time in serious meditation, before it could be stated that thermal energy was exactly the same as the energy of a moving body. In fact, a long time elapsed after Mayer's discovery before the equivalence of thermal and mechanical energy was proved. For more than half a century the doctrine of the conservation of energy had enjoyed its fixed and certain position in the Hall of Knowledge, but the kinetic theory of heat was still uncertain and it was not known whether or not, some day, it would be established, or dismissed as being an untenable hypothesis. The kinetic theory has only been recognized generally since about 1900.

2. The energy law applies not only to mechanical and thermal processes, but to all types of energy in general; its great significance is to be found in the fact that it is a universal law. It is an all comprehensive law, applicable to all energy relations, there being no phenomena in the universe to which it does not apply. If there were only one process for which it were not true it would immediately lose its position as a universal law of Nature. The validity of the energy law would gain but little if one could derive it for the case of heat phenomena directly from the mechanical energy law.

Historical. The law of energy, or as we now say, the law of the conservation of energy, as a universal law of nature, was expressed for the first time in 1842 by the physician **Julius Robert Mayer.** It is therefore not surprising to find that his thoughts were first directed to those things which had become familiar to him as a physician. He had observed that, for persons in the Tropics, the colour of the blood in their veins was not very different from that in the arteries; in cooler climes there was a marked difference—the blood in the veins is darker than elsewhere. From this he concluded that the temperature of the body must be maintained by the blood, and that in cooler lands the blood must be decomposed more so that the chemical changes which then take place can produce the heat necessary to maintain the normal body temperature. At a later date he remarked that, having had his attention drawn to this fact, he began to consider all processes involving changes in energy. It has often been noticed that Mayer's knowledge of physics was so small that he was not able to inform his contemporaries in a concise and clear manner of the things he had discovered, so that his claims did not receive attention for many years. Meanwhile **Joule** and **Helmholtz** had extended these conceptions independently of him, and had become recognized as the discoverers of this doctrine. It is quite certain that Mayer had not been able to express himself intelligently on this point, yet the fact that he had such ideas, proves quite definitely that these questions had attracted his attention. Moreover, his first publications dealing with this problem do not leave any doubt about the fact that he was quite clear concerning the importance of this law, and that it would probably be the starting point from which an exact proof would eventually come. It appears, however, and his whole life's work confirms this supposition, that his knowledge of physics was the result of intuition rather than that acquired by the study of formal proofs.

We shall now try to obtain a clear conception concerning the content of the energy law, then concerning the proofs which have been produced in evidence for it, and finally concerning the extent to which the law is true.

1. **The Conservation of Energy.** We can only deduce inferences on the properties and nature of bodies from the effects to which they give rise. The forces which reside in bodies, too, are only inferred from the effects they produce. It is true that it may happen that two oppositely directed forces may be in equilibrium, the effect of one being counterbalanced by that of the other, and it may be thought that, from the fact that two forces are in equilibrium, the means were at hand wherewith forces could be recognized or even measured. It must be remembered, however, that such knowledge is only secondary, for it always presupposes that we have learnt to recognize at least one of the forces in some other way. If we cannot recognize one of them, the fact that the whole is in equilibrium teaches us nothing, for we cannot possibly know of their existence. The fact that there is equilibrium between two forces simply means that the effect expected from the one force is exactly balanced out by that from the second. It therefore remains a fact that forces can only be recognized by the effects they produce. All forces tend to produce an effect in which some resistance is moved in a definite direction through a certain distance. The product of the force and the distance is called work, and the natural tendency of forces in a body or system of bodies is to do work, the amount of which depends upon the facilities which enable the force to move over a longer or shorter distance. The water in a weir is acted upon by gravity, but if the water lies in a low plain, it is unable to move under the influence of gravity and can therefore do no work. The water from a mountain lake can do work, not because its mass is greater than that of the water in the plain, but because it has the opportunity to move under gravity and, in so doing, overcome resistances which may lie in its path, *e.g.* it may drive a water-mill.

The capacity to perform work is thus, on the one hand, a natural tendency of a force, but it only becomes possible when the point of application of the force can move. This capacity for doing work is called energy. To each kind of energy which exists there is one element in common, *i.e.* a length, but the second element, the force, is different. Hence there are just as many types of energy as there are different kinds of forces, if we regard the inertia of a body as a force which must be overcome before the body can be moved. The different types of force and energy which are known to-day are as follows:

(*a*) The kinetic energy of inert masses.

(*b*) The energy due to gravitation, which includes potential energy. (The energy of a waterfall, the tides and the energy of the planets, in so far as their position with respect to the sun changes, belong to this category.)

(*c*) The energy due to the various types of elasticity, amongst which one has to include the energy due to sound.

(*d*) Heat, too, can cause bodies to do work, *e.g.* in heat-engines, but since we have learnt that heat is merely a manifestation of molecular energy, it is included under the heading of kinetic energy.

(*e*) The energy due to surface tension, *i.e.* to capillary forces.

(*f*) Electrostatic energy arising from the electric charges which a body may be carrying.

(*g*) Magnetic energy; the energy due to an electric current is included here.

(*h*) Chemical energy. Many substances tend to unite chemically with much violence, and can only be separated again by performing a large amount of work. Chemical energy due to combustible material, plays an important rôle in our daily life (railways, steam-engines, and cooking-stoves).

(*i*) The energy of radiation. In this connexion reference must be made to the heat and light which we receive from the sun, and to wireless waves which can encircle the earth.

(*j*) Finally, in the last few decades, the almost immeasurable energy content of the atom itself has been recognized. The seat of some of this energy lies at the surface of the atom; this energy is responsible for the radiation and light which we receive from a hot body. In some respects it can be controlled by our experimental devices. The rest of the energy has its source right inside the atom, and we have become acquainted with it from phenomena of radioactivity. We are not able to influence these processes in any way whatsoever; they can only be observed. Nevertheless the so-called disruption of the atom, which has been caused to take place within recent years, is really the first breach in a hitherto impenetrable wall [*cf.* page 163].

Later on we shall return to a discussion of these facts, which have only been mentioned here quite briefly.

Now since the capacity for doing work depends upon the magnitude of the path which can be traversed, it follows that the capacity for doing work must diminish as the path already described increases. When a body performs a certain amount of work its capacity for doing work must necessarily decrease. On examining the matter more closely we shall perceive that this diminution in its energy is not the only result of doing work, but that the amount of some other form of energy has increased. When the potential energy of a falling stone, or of a mass of water flowing downwards, is diminished, its velocity, and therefore its kinetic energy, will increase simultaneously. These laws which had so long been known in connexion with mechanical processes were extended by Mayer so that all forms of energy in general were included: **Whenever a quantity of energy of one form diminishes by any process, the same process involves an increase in some other form of energy; in fact the increase in the second form is exactly equal to the decrease in energy of the first form.** Conversely, an increase in one type of energy can only be obtained by a corresponding diminution of energy of another form.

Since this is true for all kinds of energy, it follows that energy cannot be destroyed or created by any means at our disposal. Energy can only be obtained from some other form of energy and can only pass again into some other form of energy.

Mayer expressed himself thus, and his expression is very often used even to-day: **Energy of one kind becomes converted into energy of another type.** If this expression is only used to maintain the fact that the energy which disappears is equal in amount to that which has been regained, no objection to it can be raised. But the expression promotes the idea, especially when it is a question of the **transference** of energy from one body to another, that energy is something material, which changes only its carrier or perhaps also its shape during processes on the earth. **The law of energy does not recognize this at all ;** it only asserts that **the different forms of energy are quantitatively equivalent to one another,** and the evidence in favour of the law is only concerned with this equivalence. **It tells us nothing at all about the reason for this equality.** The energy law, in any particular instance, only enables us to make conclusions concerning the energy which may have been produced, or have disappeared. The energy law states nothing at all about the process itself. That is reserved for further research. In this way the energy law has rendered service, even valuable service, concerning the nature and mode of action of the participating forces, but such could not have been inferred from the law itself.

The law of energy is sometimes expressed as follows : Let us consider any system of bodies which is shut off from the rest of the world in such a way that it neither receives nor imparts energy to external objects. The system itself can be as small or large as we like and it can contain different sources of energy inside it. It may contain a watch that has been wound up, or some other form of clock which, in running down, generates heat by friction ; or it may contain an electric battery operating a bell, driving a motor, or causing a lamp to shine. If the total energy content of such a system is considered, in which it is not possible for energy to be imparted to it or gained or lost with respect to the exterior, **the law of energy states that in a closed system the total energy is constant.** This means that if we were to measure the energy of such an enclosed system at any time—we might have found a stretched spring, some reagents capable of reacting to form other chemical compounds, and a definite amount of heat according to the temperatures of the various components—and expressed all the energy in ergs, then a definite number of ergs would represent the energy content of this system. If now, after a lapse of ten years (say), we again investigated the system and found that the spring had unwound itself, the new chemical compounds had been formed, etc., and measured the total energy content in ergs, it would be found to be the same as before, always assuming, of course, that there had been no exchange of energy between the system and its surroundings.

Quite another statement is sometimes used in order to express the law of the conservation of energy, but it means exactly the same thing however. By the term ' perpetuum mobile ', some contrivance, a machine, etc., is understood, which is capable of continuously driving some object without any energy being supplied to itself. Such a contrivance would expend energy continuously without this energy being replaced, so that it would be doing what is impossible according to the law of the conservation of energy. This law may therefore be stated : **there can be no ' perpetuum mobile '.**

2. **The Proofs on which the Law of Energy is founded.** We now pass on to consider the proofs upon which the law of the conservation of energy has been based, and perhaps this is the point at which the difference between subjective and objective reasons should be indicated. In the history of science it has frequently happened that the validity of some rule has long been known from subjective arguments before it was deduced vigorously from objective reasons. At first Mayer generalized the ideas which he had formed whilst performing medical operations, so as to include all relationships between mechanical and thermal phenomena, and in 1845

he extended them to include all phenomena in general. He found that his ideas concerning the conservation and transformation of ' forces ' (energy) could be applied to all processes and he was strongly convinced of the fact that this law was correct. Mayer thought that he had made the idea of force (energy) quite clear to himself. Actually he had only sought an idea from which the indestructibility and mutability of force (energy), which he found in all natural processes, could be uniformly expressed. In so doing, but without being conscious of the fact, he had formed the new conception of energy, which he always called force. As soon as he had recognized that this new conception embraced all phenomena, all doubt concerning the validity of his principle was banished from his mind. But this inquiring and conceptual activity must have been associated with the early period of his occupation on the subject. During his voyage he was so engrossed in this subject that he saw nothing of the glories of the Tropics ; at times he probably felt himself inspired. The one idea ever present in his mind was whether or not these ' forces ' would be conserved after taking part in some process. In 1844 he expressed himself somewhat cautiously in a letter to his friend **Griesinger** as follows : ' The day is coming when science will recognize these truths, but at present we cannot say who will champion them or when it will happen.'

From henceforward the prophet believed in the newly revealed doctrine. Mankind is for ever striving after comprehensive general truths and he finds it very attractive to possess such a law which shall be a guide for the understanding of the processes of the world. His conviction, as also his experience, that the events in the universe are regulated in accordance with comprehensive laws find in them a deep satisfaction. Many people seem to have had similar experiences to those of Mayer. We certainly hear that at first people hesitated to recognize the law but certainly not because one doubted it, but because one had not yet made oneself familiar with it. But after the appearance of the work by Helmholtz on the ' Conservation of Force ' only a few years elapsed before the law was recognized by all physicists.

But an exact science could not be satisfied with such intuitive confirmations. The law of energy is a quantitative one, maintaining the equality of the energy which disappears and that which is produced. Such an assertion is only proved when the evidence is furnished by quantitative observations.

The decisive factor was not to be found in mechanical energy, where the equality had often been derived from the laws of mechanics, as we have already mentioned. The determining factor was rather, whether a fixed quantity of some other type of energy always corresponded to the same amount of heat, or electrical energy, etc. As already mentioned, a proof will only have been obtained when heat has been obtained from mechanical energy and this heat has again been converted into an equal amount of mechanical energy. The next task is to devise some form of apparatus, which shall be as simple as possible, and which will measure all the kinds of energy which are involved. Such a task is rendered more difficult when the number of types of energy which take place in a process increases. It is therefore desirable to make an experiment in which only one kind of energy is used up, and, if possible, only one other kind appears. For example, it may be found, that x cals. of heat can be obtained from 1 erg of mechanical energy. In order to complete the proof it is not necessary to use the same apparatus to show, that 1 erg of work can be obtained from the x cals. of heat, but it is quite sufficient if this fact can be demonstrated by any other piece of apparatus. When it is finally proved that this change can take place in either direction, it only remains to find a numerical value for x which shall be as exact as possible. Any convenient apparatus can be used for this purpose, it being quite immaterial whether the heat is obtained from mechanical energy or mechanical energy is obtained from heat.

The Mechanical Equivalent of Heat. The first proof of this kind was given by **Mayer** himself in 1842, a fact which shows that he was thoroughly convinced of the importance of establishing the new law experimentally. As we have already seen, the heat which is necessary to heat 1 grm. of gas at constant volume through 1° *C.*, is less than that which is required when work is done against the external pressure. Mayer saw that the mechanical equivalent of heat could be calculated from the data furnished by these observations. Since the two specific heats were already known, and the thermal expansion of gases was likewise known, no other experiments were necessary ; he only needed to use data which were already extant. The measurements of that day were not very exact, but they have been improved ; the manner in which the calculation is made, however, remains the same. The following value is obtained [*cf.* Table 15, p. 157].

The difference between the two specific heats is

$$C_p - C = 1{\cdot}987 \textit{ cal.}$$

The gram-molecule of air occupies $22{\cdot}4 \times 10^3$ cm.^{3} and expands when its temperature is raised 1° *C.* by $22{\cdot}4 \times 10^3 \times 0{\cdot}00367 = 82{\cdot}3$ cm.^{3}. The work which must be performed against the atmospheric pressure is PV [*cf.* page 123]. Since $P = 1{\cdot}013 \times 10^6$ dynes, it follows that $PV = 1{\cdot}013 \times 82{\cdot}3 \times 10^6$ ergs $= 8{\cdot}34 \times 10^7$ ergs. These are equivalent to $1{\cdot}987$ *cal.*, so that

$$1 \text{ grm.-cal.} = 8{\cdot}34 \times 10^7/1{\cdot}987 = 4{\cdot}19 \times 10^7 \text{ ergs} = 4{\cdot}19 \text{ Joules} \quad . \quad (83)$$

This is approximately 4·2 Joules.

Only a little after Mayer, **Joule** in England made several experiments which have become classical in their fame. He began with the idea of determining this number, which has since been called the mechanical equivalent of heat, with the greatest possible accuracy, and devised new pieces of apparatus for this purpose. He caused large paddle wheels SS, Fig. 83, to be set in motion by means of the falling weights P. These wheels rotated in water, thereby causing the water to be heated. He knew the amount of energy which had been supplied, from the weight of the falling bodies and the distance through which they fell. He determined the rise in temperature with a thermometer. This, when multiplied by the weight of the water, gave the number of calories which had been developed.

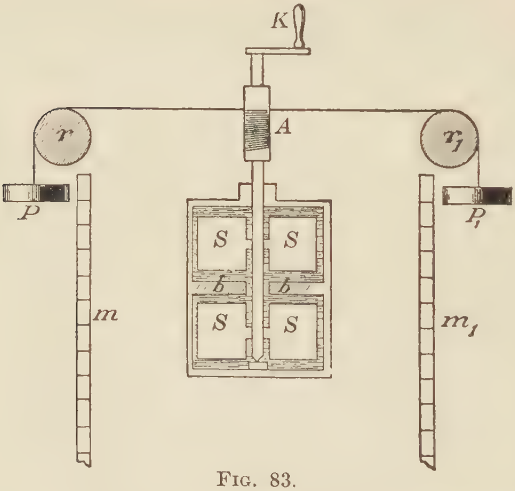


FIG. 83.

Since then, the mechanical equivalent of heat has been measured in many other ways ; there may be perhaps a dozen or more. In particular, it is very easy to use electrical methods since the ratio of mechanical energy to electrical can very easily be determined and the electrical energy can be converted into thermal energy with great ease and certainty.

Distinguished research workers have spent many years of their lives in conducting these investigations (Joule himself worked for nearly a decade —1840 to 1849). These determinations are really just so many direct proofs of the law of the conservation of energy.

The chief source of proof, however, is not to be found in such experiments. Based on these experiments the validity of the energy law has been assured and applied in all branches of physics and chemistry. These are the real proofs, for on no occasion has any fact been observed to contradict this law. These indirect proofs of the validity of the law are so numerous, various, and exact, that to-day it is no longer possible to doubt the truth of this principle.

3. **The Universality of the Law of Energy.** To what processes can the law of the conservation of energy be applied ? It can only be applied where it has been proved. It has been proved, directly or indirectly, for all processes which can be investigated quantitatively, and for nearly all forms of energy which can be measured. The most important proofs have been made in the realms of heat, mechanics, electricity and chemistry, which are the chief sources from which we can obtain energy. At the present time we have no means of measuring the amount of energy inside atoms. We can only ascertain how much energy is emitted during a radioactive change, but we cannot prove experimentally that the radiating body has lost this energy. This does not mean that we have to doubt the validity of the law of energy. For since, in all processes where measurements can be made, the legitimacy of the doctrine has been proved, we are convinced, on the contrary, that it is also valid in all other cases where we are not in a position to confirm the law experimentally.

Does this law apply to the heavenly bodies which we cannot subject to experiment ? As far as the mechanical energy which they must possess in virtue of their motions is concerned, the law can be tested : its validity has been established whenever this has been done. The remaining effects, the enormous energy which the sun's rays bring to us unceasingly, support, at least in so far as they come within the powers of observation, the doctrine of the conservation of energy. It is even so with the radiation we receive from other stars. By examining the light they emit we have learnt that the stars are composed of the same elements as are found in the earth, and the light emitted by them has exactly the same wave-length as the light from the same terrestrial elements. In those instances, therefore, which we have been able to test, especially with reference to the emission of light and the effects of gravitation, it has been shown that the processes which are at work in the heavenly bodies are governed by the same rules as are applicable on the earth. Hence, from a rational point of view, it is impossible for us to doubt that the energy principle applies to celestial bodies. If one wished to doubt it, then the investigation of the stars in general would have to be given up. For the same reason it could be required that we should limit our discussions to objects appearing on the earth's surface, since bodies inside the earth cannot be investigated.

We are not able, however, to reply in the same terms when we are asked whether the law applies **to living creatures.** Processes which occur in the animal world are apparently so fundamentally different from those in the inanimate world, that an argument which has been applied in connexion with the latter ought not to be assumed valid for living matter without further discussion. Suitable proofs must be sought for them; they will be cited in the immediate sequel. In the form of nutriment, energy is daily being supplied to our bodies. Strictly speaking food does not contain any more energy than the coal which we use to make a fire. The energy is only obtained when the coal combines with atmospheric oxygen, but since plentiful supplies of this latter substance are always available, at least on this planet, the distinction is withdrawn, because the energy actually becomes available when the coal or food has been procured.

In our bodies the food combines with the oxygen from the air which we have inhaled, and the heat of combination is the energy which is supplied by the food we have consumed. This heat of combination is first measured outside the organism by means of a bomb calorimeter. A known mass of food N, Fig. 84, is obtained in as finely divided a form as possible, and placed inside a steel bomb B, which is filled with oxygen when its lid has been screwed down. A fine platinum wire P is connected to two wires DD, and raised to a white heat momentarily by means of an electric current. The food then burns, *i.e.* it combines very rapidly, in fact almost with explosive violence, with the oxygen. The steel bomb must be rigid enough to withstand the excessive pressures which may be developed. The heat of combustion raises the temperature of the bomb. In order to measure this quantity of heat the bomb is placed in a vessel containing water so that the heat is distributed between the bomb and the water. By means of a thermometer T, placed in the water, the final steady temperature can be ascertained. If the mass of the water has already been determined and the specific heat of the steel is known, the quantity of heat necessary to raise the temperature of the whole through 1° C. can be calculated. Let this quantity which is called the thermal capacity of the whole apparatus be denoted by A. If the rise in temperature is $t_2 - t_1$, the heat developed by the reaction is $A\ (t_2 - t_1)$. If therefore a definite quantity of a substance under investigation is supplied to a living creature, the amount of energy which has been given to it is known, at least if this energy had been obtained by combustion outside the animal's body.

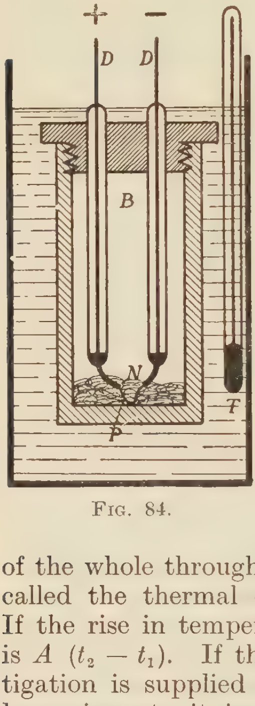


Fig. 84.

We have now to try and discover whether the same amount of energy is developed inside the body. At first this question was not attacked directly, since it appeared too difficult to measure all the energy which was being supplied to the body. The food of man consists essentially of three different classes, proteins, fats, and carbohydrates (bread, potatoes). In the first place, the quantity of each variety of substance which had to be given to the body, so that its mass neither increased nor decreased, was determined. The result was as follows: if one wishes to feed the animal at one time on fats and at another on carbohydrates, then the quantities of these which are necessary are such, that if consumed externally the amount of energy liberated would be the same on each occasion.

At a later date the question as to whether the energy produced by an animal was equal to the energy derived from the food consumed was attacked directly. An animal can produce energy in several ways. First and foremost the body temperature is always in excess of that of its surroundings so that the body is always imparting energy to them. Since, in spite of this, the body temperature remains constant, an amount of energy equal to that which is lost, must be continuously produced. Moreover, an animal can perform external work. Hence when one has succeeded in measuring the amount of energy produced by an animal, and imparted to external objects, it is possible to answer the question as to whether the energy which has been produced is equal to that which the food has supplied. For this purpose **Rubner** first built his so-called animal-calorimeter. It consists of a large glass vessel in which a dog could live for several days. The nourishment which was given to it was always of such a quantity that the mass of the animal did not change. At one time the diet consisted of pure fats, then of fats and meat, whilst finally a meat diet was given. The calorimeter had double walls and all the energy produced by the animal was transmitted to the space between the walls and there determined by measuring the rise in temperature of the air. The result of all these experiments was to show that the energy supplied by the diet was equal to the thermal energy produced by the animal to within 0·5 per cent. It was therefore proved experimentally that the animal's heat consumption, at least when its body was at rest, was exactly equal to the heat produced when its food was burnt.

In order to pursue the question further, **Atwater**, in America, built a large calorimeter in which a man could exist comfortably for several days. An analysis of the gases which were necessary for respiration was made, and their temperature measured ; the gases which were removed from the calorimeter were also analysed. A constant stream of water passed through the double walls of the calorimeter from a thermostat. After the lapse of some time the temperature of the water as it left the calorimeter became constant, *i.e.* the so-called steady state had been reached. Under these conditions the thermal energy which is imparted to the water, say in an hour, is equal to the energy coming from the man in the same time. The change in temperature of the water is small, and very difficult to measure with any degree of precision. This difficulty was overcome by performing a second experiment in which an electric current was used to heat a resistance wire placed inside the calorimeter. The current was adjusted so that the rise in temperature was equal to that which had occurred in the first experiment when the man was enclosed inside the calorimeter. Then the energy which the man produced was equal to the heat which had been developed by the passage of the electric current along the resistance. This latter quantity could be determined very exactly by means of electrical instruments, so that the first quantity thereby became known. These experiments showed that the heat produced by a man's body was equal to that which the food produced on combustion to within 0·1 per cent. Thus, within the limits of experimental error, the two amounts of energy were shown to be equal.

Afterwards the person inside the calorimeter was supplied with a machine, a tread-wheel, which he could operate by performing mechanical work upon it. The result was to show that just that excess of food was now necessary, in order for there to be no change in the person's weight, which would yield an amount of energy equal to the mechanical work done. Hence the external work, which a person does, is derived from the food which has been eaten.

These experiments refer to the lowest activity of an organism, growth, or existence. Is it thereby shown that, in general, man cannot produce energy ? In answering this question we must be careful to differentiate between things which are facts and things which are probable. Is an act of the will, of desire, or of acquiring knowledge a process involving energy changes ? In the deepest sense it is certainly something different and higher. That it is, however, made in conjunction with such a change is quite possible. The process of vision certainly begins with some effect taking place upon our eye, and this action which is caused by the light, is certainly one involving energy. Now if heat were produced in measurable amounts when these mental processes occur, and it did not arise from the food which had been consumed, then the above experiments would certainly have detected it. Actually the experiments showed that nothing takes place inside an animal's body whereby a measurable amount of energy is produced. If, however, the question is raised as to whether these experiments have proved conclusively that mental activity produces no energy, we must reply by saying that, within the limits of experimental error, all the energy developed in a person's body arises from the food which has been consumed. The error in these experiments is only about 0·1 per cent.

With plants, experiments having the same degree of accuracy have not yet been possible. The energy changes which occur during the usual life of plants are so extraordinarily small, that one cannot speak of their temperature being in excess of that of the surroundings. Moreover a large amount of the energy needed by a plant is derived from the air and the radiation it receives from the sun. It would be extraordinarily difficult to ascertain how much of this incident radiation from the sun was utilized by the plant and how much remained unabsorbed. Finally, energy transformations of reasonably measurable amount are associated with such extensive objects that it would be exceedingly difficult to accommodate them in a closed system. The few experiments which have been made with plants, however, do not make it necessary for us to suppose that the energy changes which occur during the life of a plant form an exception to the general law of the conservation of energy.

If we again attempt to summarize the known facts about the doctrine of the conservation of energy, we may do so as follows. The law of the conservation of energy is the most universal of all the laws found in Nature. There is no process whatsoever to which it does not apply exactly. It is immaterial whether the process occurs on the earth, in the sun, or on Sirius, or whether the experiment is made in the physicist's laboratory, or in the retort of the chemist, or whether the process occurs in the bowels of the earth.

Mankind has a strong propensity to give material form to such conceptions as have come to be trusted through familiar use, and gradually to allow them to develop into independent entities. Thus, in regard to energy, we have frequently read, especially in recent years, that the indestructibility of energy in all its various forms points to a unified entity. It would certainly simplify our conceptions if we could suppose that in all processes we had only to observe the transformation of an unchanged and unchangeable Something, an Energy Substance. But for this our experience so far gives us no manner of justification. In single cases, as for instance in the mechanical energy which presents itself in a vibrating tuning-fork, the acceleration is derived inevitably from the elastic force, and the conservation of energy of that system therefore follows of necessity. We are therefore not justified in drawing far-reaching conclusions from these phenomena, or even from instances in which the connexion is less easily traceable by the eye. It would undoubtedly be an achievement if we could in the course of time penetrate more deeply into this great mystery, but for that result new and convincing facts are needed ; the law of energy itself does not permit of such a conclusion.

2. THE SECOND LAW OF THERMODYNAMICS—ENTROPY

The law of the conservation of energy is supplemented by a second equally comprehensive law, and one which is more fundamental in its meaning than the first. It is generally known as the **second law of thermodynamics**, but is sometimes called the **law of entropy.** The first law tells us about the amount of the exchange which occurs when one form of energy is converted into another, but as to whether the exchange can occur or under what circumstances it does occur the first law says nothing. It merely states **that the energy which has disappeared in any process is equal in amount to the new form of energy which is produced by it.** The second law now teaches us something more, for it indicates the ease or difficulty which attends the conversion of one form of energy into another. This question would have little significance to us if it only supplied the conditions under which the exchange could take place in one or in the other direction. Actually, however, it tells us much more.

Energy Exchanges in Nature. Let us consider some of the exchanges of energy which occur in the universe. The chief source from which this earth and the remaining planets of the solar system receive their energy is the sun, which, by day and by night, and in all directions, pours forth energy unceasingly from its abundant store. In its rays, which warm the objects they encounter and sustain life, there is sent out an almost infinite supply of energy. The amount of this has been determined by means of terrestrial experiments, and especially by those which have been conducted on the tops of lofty mountains ; it has been found that every minute the earth receives $2{\cdot}5 \times 10^{18}$ cals., a quantity of heat which is sufficient to melt a cube of ice, the edges of which are almost 2 miles in length. Even if we are amazed at the extraordinarily large magnitude of this energy we have yet to remember that this is only a vanishingly small fraction of the total energy which the sun is sending out ; in fact only 1/2200 of a millionth part of the energy coming from the sun reaches this earth, and yet it is sufficient to cause Spring and Summer, and to give food and clothing to all the millions who find their homes upon this planet. What becomes of the remaining energy which the sun emits ? A few similar portions are sufficient to account for the energy which the remaining members of the solar system receive ; and the rest ?—we only know that it is lost in the vast depths of space.

Let us therefore consider that portion whose vicissitudes can henceforth be followed on this planet. It has been calculated that about one quarter of the incident radiation is absorbed in the earth's atmosphere which becomes heated a little. The remainder is incident upon the earth's surface ; part is incident upon land, but the greater part falls upon water. The portion which falls upon the land heats the rocks and the sand of the deserts so that the incident radiation is converted into thermal energy. By night the earth gives up some of the blessing which it has received during the day, for the heat is radiated into space. A very small portion falls on fields and woods, and this is in part stored as chemical energy in the plants.

The greatest portion falls on the sea, the temperature of which is raised by a few degrees, so that the incident radiation again appears as thermal energy. A smaller portion causes some of the water to evaporate, and ascend to form clouds. The potential energy which these clouds possess has been acquired from the sun. The rain falls to the ground where it forms lakes and rivers and the kinetic energy which it has obtained in falling appears as heat. Only a small fraction of its former potential energy remains latent in the water of mountain lakes. This energy can be converted into electrical energy by means of suitably installed turbines. The current is then distributed to the towns and villages in the neighbourhood. During the day it drives machinery, perhaps grinding corn to flour ; but the energy which is spent in doing this, is converted into heat. It drives the locomotives on our railways, but when they have again become stationary at their destinations, all the energy has been converted into heat. At night this electric current lights innumerable lamps, all this energy eventually appearing as heat—What has become of all the energy we have received from the sun ? In the very shortest of times it has been converted into heat, only a very small fraction remaining in the wood of trees, or in the water of mountain rivers and lakes. The wood which grew upon this planet at a very remote era is now mined as coal. When this coal is burnt some of the energy which the sun emitted ages ago appears as heat.

What happens with these larger bodies also occurs with smaller ones—even with the very smallest. When the blacksmith's heavy hammer strikes the iron, what becomes of the energy ? The iron is heated somewhat, and then begins to cool giving its energy up to objects in its immediate vicinity—A hammer strikes a bell causing it to vibrate and so disturbs the air which lies near ; the bell and the air are heated almost imperceptibly. When the bell becomes silent all its energy has been converted into heat.

Hence we see that in all the processes which occur in the world there takes place, without exception, an increase in the amount of thermal energy. The increase occurs according to the rules we have already observed ; in general there is no kind of energy, be it mechanical, electrical or chemical, which is an exception to them.

We are also forced to notice the very considerable amount of energy which the sun must possess ; if this quantity is finite then the continuous radiation which is taking place must cause the sun to cool ; the energy which was hitherto concentrated in the sun and which helped it to maintain its high temperature has been dispersed throughout the universe. Radiation from the sun will continue to take place until its temperature cannot be distinguished from that of its surroundings. We see, therefore, that all hot bodies radiate energy, which becomes converted in the shortest possible time into thermal energy at a uniform temperature. What is happening to our sun must, of course, be happening with other suns. Wherever there is a body at a high temperature it will radiate energy to its surroundings and this passes in the least possible time into heat uniformly distributed amongst the various bodies which form the surroundings. This process goes on at a very rapid pace and if, as is very probable, it will not cease for a long time yet, this is due to the fact that the sun has almost an unlimited energy content, or—can the process be taking place in the reverse direction somewhere in the universe ? Can the energy which has once been emitted ever again return to the sun ? Under such circumstances the sun would certainly maintain its high temperature, and could still send forth its energy as radiation, without ceasing, through all the ages. The second law of thermodynamics answers these questions, for it tells us that such a process cannot take place in an inverse direction, so that the celestial bodies must cool and cannot emit the same radiation twice.

Positive and Negative Processes. In order to show this we have first to divide all natural processes into two classes : in the first class there are to be found all processes which occur spontaneously, *i.e.* they occur without any external agent causing them to take place. All the conditions necessary are present in nature itself, and the process simply occurs whether we wish it or not. First and foremost must be mentioned the passage of heat from a body at a high temperature to one at a lower temperature ; it does not matter whether the transference is due to radiation, or because the bodies are in contact so that the heat is lost by conduction. All kinds of energy are continually being converted into heat and so belong to this category. Every rotating wheel rubs on its axis, and its motion will continue only until such time as all its kinetic energy has become heat, unless some external agent replenishes the energy which is being lost. The energy of an electric current which lights our lamps becomes completely converted into heat. In addition there is also another type of process which develops by itself. Suppose that we have two glass vessels, Fig. 85, connected together by means of a tap as shown ; let one sphere contain hydrogen and the other oxygen. If the tap is opened both gases will begin to mix without our aid, and after a little while the hydrogen and oxygen will be distributed equally throughout the two containers. If we were now to wait until all the hydrogen should again ‘ by chance ’ be collected in one sphere and all the oxygen in the other, our waiting would be in vain. The two gases can mix without external aid, but they cannot become separated again without it. It is just so with other processes in which mixing occurs. If two heaps of sawdust, one of oak and the other of beech wood, are placed near together and then exposed to the wind, the two heaps will soon disappear ; all the particles will be uniformly scattered and one would wait in vain for the two kinds of sawdust to collect themselves together again. This type of process, for which the necessary conditions already exist in Nature, is called a **positive process.** Should it be necessary, however, to separate again the components of a mixture, then means for doing this have to be devised. When we wish to obtain mechanical energy from the heat of steam, a machine is necessary for this purpose, and it is only about one hundred years ago since such machines have been built, whilst the conversion of mechanical energy into heat had been resorted to from very early times.

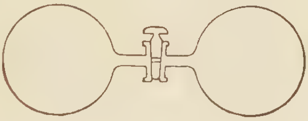


Fig. 85.

The various types of energy can be divided into three classes, according to the ease with which they can be used and converted into other types of energy. The *first* class comprises all those types of energy which, without exception, can be converted into all other forms. This is the most useful sort of energy and is found as the kinetic energy of moving matter, the potential energy due to gravitation, electrostatic and electrodynamic energy, and providing the elastic limit has not been passed, it is to be found as elastic energy. Even in such processes as these all the existing energy is not converted into useful energy of the same kind. In every movement frictional losses occur at the surfaces of contact and also in moving against the air; this energy is converted into thermal energy which is less useful. By carefully designing the apparatus, however, this loss of useful energy can be made very small. As far as the celestial bodies are concerned, a loss of such useful energy as they move along through space has not yet been proved with certainty, but for all motions taking place on our earth, or in the air which surrounds it, this loss of energy cannot be avoided.

The second class into which some types of energy fall is such that they can be converted partly into useful energy but the remaining portion must inevitably become useless. In this category is to be found the thermal energy of a steam boiler, in fact all thermal energy belonging to bodies at a higher temperature than their surroundings. Likewise chemical energy finds its place in this class. In all chemical processes a change of energy takes place; very often in large amounts (combustion of coal, explosions, etc.), whilst in other cases it may be small or even only just detectable. When such a change occurs, a definite fraction of the energy is converted into mechanical energy or some other form of energy belonging to the first class, *e.g.* we may cite the electrical energy of an accumulator, etc. The amount of this energy, if the process occurs at constant temperature, is called, according to **Helmholtz**, the **free-energy** of the process. The energy possessed by the sun's radiation also belongs to this class. The remaining portion of the energy always becomes converted into thermal energy.

Finally, the third class of energy comprises the thermal energy of bodies in a large extended space, in which the various bodies have no temperature difference with respect to one another. This is the condition which finally occurs when a hot body is allowed to cool by imparting radiant energy to its surroundings, or when a moving body is allowed to continue to move until all its energy has been spent in overcoming friction. This last kind of energy is the least useful of all, for it is impossible even to obtain any portion of it in the form of useful energy.

Every event which happens in the material world can therefore be placed in one of two large classes according as the energy can pass from a higher type into one of the remaining classes, *i.e.* from class one to two, or from class two to three, or whether it can change from three into two and from thence into one. The first group is said to contain processes which are **positive**; the others which occur in the reverse direction are called **negative.**

The second law of thermodynamics states that positive process can occur spontaneously, *i.e.* without external aid, but, in general, processes which are negative in character are not possible unless they are accompanied at the same time by some other process which is positive. **Clausius**, who first noticed these relationships, discovered that this infinitely great series of phenomena can be traced back to and epitomized in a statement which he expressed thus : **Heat cannot on its own account pass from a cold body to one which is hotter.** The words ‘ on its own account ’ simply mean that the passage of heat from a cold to a hot body cannot be a single process. It must be accompanied by some other process which aids and facilitates it. This accessory process must also be one which can occur ‘ on its own account’, *i.e.* it must take place in a positive direction.

Thomson (Lord **Kelvin**) (1824–1907) expressed this behaviour of thermal energy in the following form : it is impossible to devise a machine which will run continuously by abstracting energy from its environment. Such a machine would not violate the first law of thermodynamics, since it would not be obtaining energy from nothing, but, for example, from the vast ocean of water across which such a machine may be considered to drive a ship. Nevertheless it would be a kind of perpetual motion since only the thermal energy of the ubiquitous air, or sea, would be utilized. The sea in the vicinity of the ship would be cooled somewhat and this would supply the ship with kinetic energy. Finally, however, this energy, which had been extracted from the sea, would not cause the sea to be cooler, since during the motion of the ship its energy would become converted into thermal energy by overcoming the resistance which the water offers to its movement. At the conclusion of the voyage the small amount of energy which might still be retained in the ship would very quickly be dissipated by friction into heat. Such a contrivance would therefore permit the thermal energy of the ocean to do work literally without limit.

It is even possible to cool a body below the temperature of its surroundings, whereby its energy is given to the surroundings at a higher temperature. Such an exchange of energy does take place in an ice-producing plant, but it does not occur alone, *i.e.* it must be ‘ assisted ’ by a simultaneous process taking place in the positive direction. The above-mentioned but impossible machine may be said to produce perpetual motion of the second type so that the second law of thermodynamics can be expressed in a short sentence as follows : **Perpetual motion of the second kind is impossible.**

The statements of both Clausius and Kelvin are exactly equivalent in their meaning.

The Entropy Function. In order to reveal the truth embodied in these statements we must now discuss briefly a piece of apparatus by which work can be obtained from heat, viz. the steam engine. For the most part we shall only repeat the ideas which were first expressed by the young French engineer **Sadi Carnot** (1796–1832). He took the steam engine, which, even in those days, was quite well known as an example of a heat engine in general, and investigated the best conditions under which mechanical energy could be obtained from thermal energy. Although he regarded heat as a substance, his conclusions are quite free from this error.

Now in every actual machine friction must occur between all the moving parts, and moreover, there are always some parts of it at a temperature different from that of its surroundings. As soon as ever the machine begins to move the first condition causes useful energy to be spent in overcoming friction so that both these circumstances tend to lower the temperature differences which existed. At present, however, we shall disregard them, since these losses can be made very small in suitably arranged experiments. On the contrary we shall consider those relationships which are intimately connected with the process, and which are inseparable from it.

In an earlier chapter it has been seen that steam can only be obtained at a high pressure if its temperature is raised, for otherwise, the increase in pressure would immediately cause the steam to condense. It is therefore essential to have in every steam engine a vessel containing steam at a high temperature. This steam is allowed to pass into a cylinder, which is closed at one end by means of a moveable piston. Moreover we must have a second receptacle, which may even be the atmosphere itself. In such an engine the steam, which has entered the cylinder, can only cause the piston to move outwards until the pressure inside the cylinder is equal to that of the atmosphere, for then the pressure on both sides of the piston is the same. If the machine is to continue to do work the vapour inside the cylinder must be ejected and a new supply of steam allowed to enter. The spent steam under a pressure of one atmosphere and therefore at 100° C. must be ejected into the air, or more generally into some heat receiver (or sink). It is essential for this steam, which has done its work, to be removed, for otherwise the engine will cease to operate. When therefore the quantity of heat which is taken from the source is Q_1 only a part of this is used in moving the piston and thereby performing external work ; the remainder Q_2 will pass during the same process into the receiver as heat at a lower temperature (100° in the above example). The production of mechanical energy from the thermal energy of the source is a negative process, for it could not occur by itself ; a process in the positive direction must take place at the same time, *i.e.* heat must pass from the source at higher temperature to the sink at lower temperature. This was the ‘ compensation ’ desired by **Clausius.** He defined the magnitude of this compensation by introducing a quantity ϕ, which he called the **entropy** of the body. The entropy of a body increases as the heat contained in it increases, but decreases when the temperature rises. He therefore wrote $\phi = Q/T$, where T is measured on the absolute scale of temperature. If therefore the source of heat supplies a quantity of heat energy Q_1 at temperature T_1, its entropy decreases by an amount Q_1/T_1 ; if a quantity Q_2 is then given to the sink at temperature T_2 its entropy will increase by an amount Q_2/T_2.

It is therefore easy to see at once that all the defects in a machine must cause the entropy to increase. If, for example, the piston does not move freely in its cylinder, but performs work in overcoming friction, a part of the energy which it has acquired will be converted into heat at a lower temperature. If this fraction is denoted by q, then the second receptacle will receive an additional amount of entropy q/T_2. Hence, whenever frictional forces have to be overcome, the total entropy of the system will increase. Moreover, it may still be possible for some heat to pass directly from the source to the sink owing to some deficiency in the thermal insulation of the system. Such heat will pass from a temperature T_1 to T_2. If this quantity is q the source will lose an amount of entropy equal to q/T_1 whilst the sink will gain q/T_2. Since $T_1 > T_2$ it follows that the resultant effect will be an increase in the entropy of the system. This loss of energy can be avoided in two ways. In the first way the quantity q is made zero, *i.e.* the escape of heat from one body to another is prevented. This is done by filling the space between the two bodies with a thermal insulator ; the process is then said to occur **adiabatically.** In the second way T_2 is made equal to T_1 so that the gain in entropy of the sink is equal to the loss in entropy of the source, so that the total change in entropy is zero. If therefore the passage of heat from the source to the engine is to take place without increase in entropy, the temperature of the engine must not be less than that of the boiler. The process is then said to occur **isothermally.** When the loss of useful energy in all these processes is avoided, it is possible to get the maximum amount of work out of them which, in general, can be obtained. It is evident that these conditions can never be realized in practice, but this ideal state can be approached by reducing all possible losses to a minimum. Clausius proved that when the maximum amount of work has been obtained in this way the compensating process, *i.e.* the giving up of heat to the receiver at lower temperature, is quite independent of the properties of the working substance, and that the gain in entropy of the sink is equal to the loss of entropy which occurs at the sink. We therefore have

$$\frac{Q_1}{T_1} = \frac{Q_2}{T_2} \quad . \quad . \quad . \quad . \quad . \quad (84)$$

In an ideal process, in which all losses are avoided, the entropy remains constant, but under no conditions whatsoever is it possible for the entropy of a system to decrease. In all actual processes, where some thermal energy is always lost or spent in overcoming frictional resistances, the entropy must increase. From equation (84) follows the maximum amount of work which can be obtained when the engine is working between temperatures T_1 and T_2. The thermal energy which is so transformed into mechanical energy is equal to $Q_1 - Q_2$, and from the above equation we have

$$\frac{Q_1 - Q_2}{Q_1} = \frac{T_1 - T_2}{T_1} \quad . \quad . \quad . \quad . \quad (85)$$

This equation shows that the fraction of energy which can be converted into mechanical energy under ideal conditions only depends upon the temperature difference which is maintained between the source and the sink.

A steam engine, for example, working at 10 atmos. pressure will have a source whose temperature is 180° *C.*, or $T_1 = 453°\ K$. If it works without condensation, *i.e.* the steam merely passes into the atmosphere, the lower pressure is one atmosphere, so that $T_2 = 373°\ K$. The maximum amount of the energy expressed as a percentage which can be converted into useful work is therefore

$$\frac{Q_1 - Q_2}{Q_1} = \frac{80}{453} = \frac{1}{5{\cdot}7} = 17{\cdot}5 \text{ per cent.}$$

The remaining 82·5 per cent. merely passes at once into heat at the lower temperature.

If the steam, after having done its work, is not passed directly into the air, but is first condensed to water, at a lower temperature of about 40° *C.* by means of a supply of cold water, the above fraction, which is called the **efficiency** of the machine, becomes 140/453 = 31 per cent.

From such numbers as these, it appears that the great tendency in Nature, whereby all energy is transformed into heat at a fairly uniform temperature, cannot be counterbalanced by processes taking place in the reverse direction, for these latter are relatively few in number and exceptionally weak, when compared with the large amounts of energy associated with processes in the positive direction. Clausius has embodied the results of his researches in a short statement as follows. In every closed system which neither receives energy from without nor loses energy, the entropy tends to a maximum. This maximum is reached when all other forms of energy have disappeared in favour of heat energy at a uniform temperature. In so far as the whole universe can be considered as such a closed system of finite extent the law of **Clausius** may be expressed thus : '**The entropy of the universe tends to a maximum,**' a statement which means that the available energy of the universe is tending to zero.

It is perhaps not unprofitable to make a few remarks concerning the meaning of the entropy function. In formulating equation (84), do we imply that the second law of thermodynamics has been proved mathematically ? Or, what is really the meaning of this formula ? The rule of Clausius is merely an exact expression of the context of the law. This assertion has been of the greatest value in scientific work, and it still is, because it permits further conclusions to be expressed exactly. It has also been of the utmost importance in establishing the law since, in many instances, it was possible to test experimentally the conclusions which had been drawn. The formula itself, however, is no proof that the entropy law is valid. It has enabled the law to be used as a working hypothesis and has expressed the content of the law in an exact form, but the proofs of the law must be sought elsewhere. The proof of this law is to be found in the fact that, wherever the law has been applied in practice, the numerous conclusions which have been made with the help of the exact formula expressed above, are nowhere found to contradict it. The laws of energy and entropy are the two pillars of an extensive field of inquiry known as **thermodynamics,** in which the contents of both laws have been subjected to many various and extensive tests ; scarcely any other scientific law has been subjected to such critical examination.

To **W. Thomson** (**Kelvin**) we are indebted for a very important application of the law of entropy. He used the above equations to construct a scale of temperature which is entirely independent of the working material ; other scales of temperature based upon liquid-in-glass or gas thermometers depend upon a change in volume or change in pressure of the thermometric substance. Kelvin considered an ideal engine which worked without friction and without gaining or losing heat. If it is assumed that a quantity of heat Q_1, at temperature T_1, is taken from the source and used to do work, after which the remaining heat Q_2 is given to a sink at temperature T_2, we have

$$\frac{Q_1}{Q_2} = \frac{T_1}{T_2}$$

It then only remains to fix the unit of temperature arbitrarily, *e.g.* we may say that between the temperature of melting ice and that of water boiling under a pressure of one atmosphere there shall be an interval of one hundred degrees, in order to define a system of temperature measurement which is independent of all the properties of the substance used for that purpose.

Boltzmann's Explanation of the Law of Entropy. In spite of all the confirmations of the laws of entropy which have been obtained experimentally, most persons desire the principle to be explained somewhat more fully. We might wish to obtain an insight into the action of the forces which finally lead to such significant laws.

An important step in this direction was made by **Boltzmann** in 1868. He compared thermal movements with the other types of motion and wondered why heat energy is so very different from all other forms of energy. He came to the following conclusion. Let us direct our considerations to the experiment depicted in Fig. 85 [*cf.* page 187]. In one vessel we have hydrogen, whilst oxygen is present in the other. After the lapse of a little time both gases have become thoroughly mixed. Suppose now someone should say that he would wait until both gases were distributed again as they were initially, *i.e.* until only hydrogen was present in the one vessel and oxygen in the other, and that he would then close the tap quickly so that both gases were again separated. Can such a thing be accomplished ? Is not this condition just as possible as every other ? Boltzmann replies as follows : the condition of a perfect separation taking place is possible just like every other condition ; but the probability that this condition should arise now, or even that it should occur even in a thousand years, is so extremely small that the prospect of living to see it is practically negligible. Boltzmann applied the rules of the theory of probability, saying that some natural processes are more probable than others.

So long as there is an excess of oxygen in the right-hand container it is more probable that a molecule passing from the right-hand side through the tap will be an oxygen particle than one of hydrogen. As this process continues, however, the two gases gradually become more intimately mixed. It is only when there is the same number of molecules of each kind present in each vessel that the probability of an oxygen atom passing across is equal to the probability with which a hydrogen atom may be expected to pass, *i.e.* when the two gases have once been mixed thoroughly, they will not pass back into the original condition on their own account.

Thermal movements are only to be differentiated from all other movements by the fact that they occur perfectly at random. Such a chaotic motion must always occur of its own accord. ‘ Of its own accord ’ does not imply that it does so without cause. Everything that does happen requires a cause, in fact a cause which results in the event taking place in the manner it actually does. ‘ Of its own accord ’ would rather mean that the causes which are necessary for thermal motions to increase are always present, and it is no longer necessary to supply some original cause, for it is there already. The departure from this chaos, however, *i.e.* the establishment of order, requires some directing influence which, in general, is not present and so must first be supplied from without.

When a bullet leaves a rifle all the particles in the bullet have the same direction and velocity ; its motion is quite regular. But when the bullet is moving through the air, the molecules of air are thrust aside to the right and to the left, and behind the bullet the molecules move in eddies, to fill the rarefied space which has been created by the passage of the bullet. In this way kinetic energy passes continuously from the bullet to the air and in so doing the air molecules acquire different velocities and directions, so that the kinetic energy becomes converted into energy of a chaotic motion, *i.e.* thermal energy. Afterwards the bullet passes into a heap of sand. Since the sand particles are scattered in all directions, each rubbing against its neighbour, the remaining portion of the kinetic energy of the bullet is converted into heat. This process always takes place and quite by itself. ‘ By itself ’ simply means that the conditions necessary for the process to occur are already there, having been supplied by Nature : the air-filled space, the smallness of the air molecules and of the sand particles which can be set in motion in all directions. Nothing further is necessary for the kinetic energy to become transformed into chaotic energy of motion.

It is otherwise, however, if we should desire the process to occur in the reverse direction, *i.e.* that the heat should be obtained from the sand and air in order to drive the shot back again into the barrel of the rifle. In order for such a thing to happen it would be necessary for all the sand particles to come together at the same time and place, and to have moved in the correct direction so that they could thereby cause the bullet to begin to move backwards. If the bullet began to move through the air, then in order for its velocity to increase, all the air molecules would have to arrive without exception at the right instant with the correct velocity ; if this were so the bullet would eventually return with its initial velocity to the rifle. The process would in no way have violated the law of the conservation of energy. The energy would still exist in its entire amount and always will do, but nevertheless, the process will not occur. It requires that guiding influence which would just cause all the millions and millions of sand and air molecules to move with exactly the right velocity and appear at the right place, not a thousandth of a second too late or too soon, when the bullet passed by. Such a guiding influence does not occur on its own account in Nature. That is why the passage from an orderly to a chaotic motion always takes place ; the inverse process can never occur voluntarily.

The whole conception of Boltzmann is in no way limited to physical laws and processes. It is quite general. When we stand in the square of some city and observe its business life we see men hurrying to and fro. If we should attempt to count the number of persons passing over a bridge in each direction we should find, in the course of a day, that the numbers were very nearly the same. This does not mean that the individual men pass over in one or other direction without a reason ; no, each one has his business and that is the reason why he is passing over the bridge *just at this moment in a particular direction.* But the reasons why the people in a city are moving are so numerous and various that finally, in the course of a day, the number moving in one direction or the other is the same. If we look at this bustle as the people move in this direction or in that, we are not surprised and do not inquire why they are moving so irregularly. But when suddenly we see all the people running in one direction, we exclaim with others, 'What has happened ?' for we seek to discern some guiding influence. In other words, the reasons for a perfectly irregular motion are always present ; it is so because of the great variety of circumstances which cause one person to move in this, and another in that direction.

Similar conditions maintain when one is dealing with inanimate things which move according to fixed laws. When we leave our home for an eight-day tour, on our return we find the top of the table covered with dust. If we observe the dust particles under a microscope we find the particles distributed irregularly over the table, but at no point do we find a large space free from dust. What has put each particle in its position ? The influences which have been at work are many and various. First of all there is the place at which each particle had its origin ; then its size and shape which determine the rate at which it fell ; then the small movements in the air of the room itself which cause the particles to deviate from their rectilinear paths. In the variety which exists amongst all these causes is to be found the reason which is sufficient to account for the distribution of the particles over the whole table, and no one would sit down for a moment and wonder why they could have been distributed with such wonderful regularity over the whole table. On the other hand, if we were to find our names written in the dust we should have cause to wonder. Or perhaps we should not wonder, but be convinced that some good friend had found a key and been in the room prior to our return, for such a guiding influence as would cause the falling dust to leave uncovered the spaces which form the letters in our name does not occur on its own accord. In daily life we are all convinced of the truth of Boltzmann's assertion that a chaotic arrangement can occur on its own, but an orderly arrangement requires some guiding influence.

The Entropy Law as a Statistical Law. Having become convinced of this assertion, the whole law in its deepest meaning requires amendment. If it has hitherto been regarded as a law which is strictly true, this attitude can no longer be maintained. Now the law is found to be the result of a thousand, or of a million 'chances', amongst which, when considered in detail, not one has anything to do with the law. The law of entropy has become a **statistical law**, *i.e.* a law of very large numbers. It therefore follows that this law is only valid when we are dealing with large numbers, for it is only by having a large number of things that these accidental influences can be so large that ultimately there is no regularity at all to be recognized. The law therefore becomes a law of quite a different type—it is a **law of probability.** The probability that a certain event will follow this law will become more likely the greater is the number of particles participating in the process. If in the two vessels shown in Fig. 85, page 187, there were only two molecules, then when the tap were opened it would be very easy to intercept both molecules whilst they were in one vessel, by closing the tap at some arbitrary time; even with three or four molecules a similar condition would be possible, but we should have to consider ourselves very fortunate if we obtained it. Whether or not it could be obtained in a year if we had ten molecules is very questionable. With a hundred molecules no one would lay even the smallest wager that they could all be obtained in one of the vessels, and with a thousand molecules everybody would say it was impossible for all the molecules to be in the one vessel at the same time. But if we now apply the same argument not only to millions of molecules, but to all the molecules found in some large vessel, which may be the earth itself, it is practically the same whether the entropy law is treated as one of statistics, or as one which is absolute.

Entropy and Probability. **Boltzmann** has shown how the probability of a certain condition can be calculated. For the sake of simplicity we will assume that there are ten balls in the right-hand vessel of Fig. 85. After opening the stop-cock one or more of the balls will soon find their way across into the left-hand vessel, and finally a state will be reached in which both vessels contain practically the same number of balls. Suppose now that the stop-cock was opened and closed a thousand times and the number of balls in each vessel ascertained every time; in by far the most cases the balls would be so distributed that there were five in each vessel, but there would be instances in which deviations from this were found. Finally, the following results may be found if the experiment were repeated a thousand times.

TABLE 22

Distribution of the Spheres.	No. of Times Occurring.
5 : 5	580
6 : 4	350
7 : 3	60
8 : 2	9
9 : 1	1

We say that the various states of distribution are more probable, the more frequently they occur. The probability of the first state is 580 : 1000, whilst that of the last is 1 : 1000. From whence come these different probabilities ? Let us assume that there are nine balls in the right-hand vessel, and one in the left. All move irregularly, but on some occasions each one can succeed in passing straight through the opening in the tap. If one of the nine balls on the right-hand side enters the tap the condition in which there is a uniform distribution of spheres throughout the two vessels will have been more nearly approached ; should, however, the one ball in the vessel on the left-hand side pass through the stop-cock, then the state in which a uniform distribution prevails will be more remote than ever. There are nine balls, however, which satisfy the first condition, but only one which satisfies the second. The probability that the state of uniform distribution will be approached is therefore nine times as great as the tendency in the other direction. The state of uniform distribution is therefore called the condition of maximum probability. If our system (consisting of 10 balls in one vessel, and none in another) is left to itself, the state of uniform distribution is most probable. This is really only another way of stating Clausius' expression, that the entropy of a system always tends to a maximum. When Boltzmann succeeded in calculating the probability with which a condition is likely to exist in a system containing many molecules, he found that the entropy, ϕ, could be expressed as the logarithm of the probability, W, *i.e.*

$$\phi = c \log W,$$

where c is a constant.

The Ultimate Thermal End of the Universe. The law of entropy shows that at some future time the universe will arrive at that condition in which no more useful energy will be available. Several physicists have thought that they have seen a method whereby this catastrophe could be avoided, for since the law of entropy is a statistical one, it does not preclude the possibility of some spent energy collecting together and again becoming useful.

The author is unable to accept this view. If at any time a pin's head should, on its own accord, become twenty degrees hotter than its surroundings, we should at once seek the agent which was responsible for this change ; but no physicist would dare to assert that the heat energy had by chance been acquired by the pin's head, and the temperature of the surroundings thereby lowered. But the assumption that by similar processes a body of the size of the sun should gain a temperature ' fortuitously ', such that it would become capable of radiating so much energy that the other heavenly bodies at their large distances would become appreciably heated by it—this assumption is a perfectly wild one.

3. THE THIRD LAW OF THERMODYNAMICS

In 1906 **Nernst** presented to us the third law of thermodynamics, which, like the two first laws, arose out of a rich store of observed facts; it governs extensively naturally occurring processes, and can be expressed mathematically in the form

$$\lim_{T \to 0} \frac{dA}{dT} = 0.$$

It has become one of the most important laws in thermodynamics. Its chief importance lies in its application to physical chemistry, the field of knowledge in which it originated. It encroaches, however, upon the domain of pure physics, and numerous deductions have been made by means of it; some of these will now be discussed very briefly.

The starting point of the considerations which finally led Nernst to recognize this law arose from a series of different observations at low temperatures. In discussing the specific heats of bodies we have already pointed out that they all decrease at low temperatures; the specific heats of solid bodies decrease so much at the lowest temperatures that can be obtained, that Nernst formulated the hypothesis that at the absolute zero the specific heats must become zero. In fact, many substances, before they have reached this point, give up no more heat even when they are cooled still more severely. Further investigations finally led to the result that all those properties of a body which varied with the temperature at higher temperatures, would be independent of the temperature in the neighbourhood of the absolute zero. The volume of a gas at constant pressure will no longer decrease when the temperature is lowered; likewise the pressure of a gas at constant volume will not vary with the temperature. The resistance of a lead wire towards an electric current decreases very suddenly at very low temperatures; **Kamerlingh-Onnes** recently found that it amounts to only 10^{-12} of its value at 0°, so that enormous currents can be passed along the wire without heating it appreciably. The thermoelectric force [*cf.* page 253] likewise becomes zero. The so-called paramagnetic bodies acquire, at ordinary temperatures, magnetic properties when they are placed in a magnetic field; but if they are in the neighbourhood of the absolute zero their susceptibility becomes very small and approaches a zero value.

In order to be able to point out, in some measure, the connexion between these relationships, we must pursue the subject somewhat more deeply.

Every system of bodies has in every condition (determined by its temperature, pressure, and volume) a definite energy content, which depends especially upon its temperature—let it be called U_1. Now let the system change its state so that its energy content becomes U_2; the difference in energy is $U_2 - U_1$. The energy law tells us that this difference in energy is independent of the way by which the system passed from the state U_1 into the state U_2. If it were not so, but there were two paths along which the difference in energy varied, we could let the system pass along such a path from (1) to (2) that it gained the most energy and then let it return along a path from (2) to (1) over which the least amount of energy was given out. If we had obtained energy by such a process it could be used to do useful work continuously, *i.e.* work would be obtained from nothing. It is to be remembered that U is a symbol for all the energy of the bodies which participate in a process. Since it is easy to convert all kinds of energy into thermal energy, the total energy can readily be measured. When no special precautions are taken the total energy will of itself be converted into thermal energy. We therefore only need to obtain some apparatus which will allow us to measure the thermal energy produced. It is only necessary to allow the process to take place in a closed vessel having double walls between which a stream of water is passed. All the heat produced will be given to this water the rise in temperature of which, as measured by a thermometer, enables the heat energy to be calculated.

Let us further assume that this process takes place at two different temperatures ; let ΔT be the change in temperature which corresponds to a change in the energy content which we call ΔU.

The ratio $\frac{\Delta U}{\Delta T}$ is not the same for all temperatures. It can only be considered constant for small differences of temperature when it is written in the form dU/dT. Nernst's law then states : **In the neighbourhood of the absolute zero the change in the energy content, which is ascertained from experiments at different temperatures, is equal to zero.** In symbols we have

$$\lim_{T \to 0} \frac{dU}{dT} = 0.$$

In the above the total change of energy has been expressed as heat energy. That serves to define the quantity of energy which has been converted from one form into another. For a process to take place, however, *e.g.* a chemical reaction, another factor is of critical importance ; what is the amount of the energy set free by the process which can be converted into useful energy : and in particular, what is the maximum value of useful energy A which can be obtained from the process. In a former chapter we learnt from the law of entropy, applied to the special case of heat being transferred from a hot body to a cold one, which is essentially what occurs in a steam engine, how to calculate the value of this useful energy. The amount of available energy was given by the expression $Q_1 - Q_2$ of equation (85). If, in this equation, we put $T_1 - T_2 = \Delta T$ and the available energy $Q_1 - Q_2 = \Delta A$, we have

$$\Delta A = \frac{Q_1}{T_1} \times \Delta T.$$

The greatest value of the available energy is likewise dependent upon the temperature. In all processes there exists a maximum value of available energy which can afterwards be converted into all other kinds of energy. **Helmholtz** called this the **free energy.** In determining experimentally the maximum value of the free energy, which can be obtained from any process, it is essential to avoid all energy losses due to friction, or the passage of heat to bodies at lower temperatures. If the energy contained in a system can be converted into all other forms of energy, it is also necessary that one should be able to convert it into that form of energy from which it originated. In other words, the maximum amount of free energy is only obtained when the process can return to its original state by utilizing the same quantity of energy $A_1 - A_2$. The process is then said to be **reversible.**

The most useful processes, therefore, would be those in which all the energy which was exchanged would appear as free energy. In practice, there are many processes in which this condition is very approximately realized. Electric batteries are such that chemical energy is converted into electrical energy. In many cases these processes are very nearly reversible, *i.e.* if the same amount of energy is returned to the battery the same amount of chemical energy is obtained again. The lead accumulator very nearly belongs to this category. In using it, the lead and lead peroxide of the electrodes are converted into lead sulphate; in the so-called process of charging an accumulator electrical energy is supplied, and chemical energy given to the electrodes. If all losses caused by using too strong a current are avoided, the electrodes are completely reformed by supplying an amount of energy equal to that which has previously been taken from the accumulator during the time it was being discharged. In other instances the difference is not very great.

The amount of free energy which can be obtained in any process depends upon the temperature. The maximum value which it can have would be $A = U$, say. This would mean that all the energy given out would be available; a greater amount is excluded by the law of the conservation of energy. Now **Nernst** noticed that A approximates to this maximum value when the whole process occurs at a very low temperature, and he concluded that in the neighbourhood of the absolute zero $U - A = 0$.

But since at these lowest temperatures the change of U with temperature is also equal to zero, we have

$$\lim_{T \to 0} \frac{dU}{dT} = 0,$$

so that the same must occur with A, for otherwise it could not be equal to U in the whole region around the temperature of absolute zero. The law which is valid must therefore be capable of being represented by

$$\lim_{T \to 0} \frac{dU}{dT} = \lim_{T \to 0} \frac{dA}{dT} = 0 \quad . \quad . \quad . \quad (86)$$

If the variations of U and A are represented graphically, the fact that $\frac{dU}{dT} = 0$ signifies that the curve shall be horizontal when $T = 0$; moreover, since $U = A$, it follows that both curves must touch one another when $T = 0$. For other values of T the value of U must always be greater than that of A. The type of curve which is therefore obtained is shown in Fig. 86.

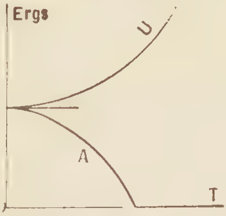


FIG. 86.

This non-variation of the energy content with temperature in the neighbourhood of the absolute zero is in complete accord with the observations which were mentioned above. It has led to the deduction of a very important conclusion in theoretical science to which we shall shortly refer. Nernst has used it to express his law more tersely: **It is impossible to reach the absolute zero of temperature.** Worded in this manner it takes its place along with the two other laws of thermodynamics, which state the impossibility of producing perpetual motion of the first and second kinds respectively.

In order to show somewhat more closely the connexion between this mode of expressing the law, and that adopted above, let us now consider by what means one can endeavour to approach the absolute zero. We shall find that all of them have to be put aside on account of the behaviour of bodies at low temperatures as previously explained. The chief method, which is the one used in **Linde's** liquefaction plant, is one in which a gas is considerably compressed and then suddenly allowed to expand. During the expansion, the molecules of the gas have to do work as they seek to move further apart from one another and against the forces of molecular attraction. This work is performed by the gas at the expense of its thermal energy. When, however, the pressure can no longer change with the temperature, a change in pressure can no longer produce a change in temperature. Since similar things happen with all properties of matter depending upon the temperature, it follows that it is impossible to reach the absolute zero.

It would be quite wrong, however, to conclude from this that important information concerning the behaviour of bodies at the zero of absolute temperature must for ever remain hidden from us. It is rather the contrary which is true, for the absolute zero is no longer able to withhold its secrets from our ken. The behaviour of a body has ceased to vary long before the absolute zero is reached—one might say it has become 'frozen in'—so that no further changes are to be expected. When therefore it is not a question of attaining this zero, which is at $-273{\cdot}2°\ C.$ as measured by a gas thermometer, but merely of discussing the properties of substances at this temperature, it can be quite rightly maintained that the substance has acquired the properties it would have at the absolute zero even before the absolute zero has been reached.

The fact should not be concealed that the foregoing suggestions do not represent at all fully the main content and importance of this law. The importance of the law is greatest in the fields of chemistry and physical chemistry where it has been much used. In particular it will decide whether some change will proceed in one direction or the other, or not at all. Moreover, discussions concerning the meaning and significance of the law are still taking place, so that even on this ground alone, a final judgment cannot yet be expressed.

CHAPTER IX

THE ATOM

1. The Existence of Atoms

THERE was a fundamental difference between the atomic theory of **Democritus** and his pupils and that of **Dalton** and the flourishing schools of chemistry of the last century. The difference lay essentially in the foundation of the theory. The ancient Greeks inferred their teaching from observations which were few in number and generally based on phenomena which occurred spontaneously ; in general the phenomena which they interpreted were sometimes used by another school of thought to support an entirely different theory. Since their theories were not founded on experiment they were not conducive to the design of new experiments. Their teachers were thinkers, but not researchers.

Things are quite different with the modern atomic theory. With a small number of chemical compounds the law of constant proportions and that of multiple proportions had been found to be true. On these facts **Dalton** based his theory. He saw that these relationships were necessary consequences, if the smallest particles of the substance under investigation were all equal in mass. Since the theory arose out of a field in which research was very active, it soon led to new questions and answers : the discovery of elements, the determination of their atomic weights, the analyses of compounds, the construction of chemical formulae, the laws of synthesis and analysis, and in general their physical and chemical properties. Out of Dalton's theory our chemical laboratories have sprung up, and from them our large chemical works have originated. All these consequences are really just as many new confirmations of the atomic theory, through which alone all these phenomena have become intelligible. It is therefore not surprising to find that chemistry unanimously recognized the atomic theory ; adherents to any other theory are no longer to be found. Even if **Ostwald** stood aside from all his contemporaries at one time, his later conversion to the atomic theory brought all the more support in its favour.

The assumption that the whole of matter is made up of atoms is quite generally thought to apply to all bodies and all states of aggregation. If it were a conception, as already mentioned once before, which wished to deprive some states of aggregation from having an atomic structure, it would have to be discarded completely, since it would necessitate an intimate change in the atom as it passed from one state to another ; there is no evidence at all in favour of such a proposal. It is sufficient, therefore, if we could succeed in proving that atoms do exist in any one form of matter. Facts in chemistry, which have already been mentioned, have given a solid foundation to the idea of atomic structure. The persuasive power of facts, however, is so great that a hypothesis which includes them ceases to be a hypothesis. Although, on subjective grounds, we are convinced that the atomic theory is true, the wish still remains for some process to be found in which the atomic structure of bodies, as such, would reveal itself to us. Such observations would make a hypothesis into a fact, and they would be the last and final confirmation of something about which we had previously only an indication: they would crown the work of human minds. For a long time it appeared that our work would never be able to bear such a seal.

All experiments, which were made in order to obtain an estimation of the size of an atom, indicated such extraordinarily small values that one hardly dared to hope that these very tiny quantities of matter could ever be indicated, whether by the balance, the eye or the sense of smell.

Layers of Minimum Thickness. Monomolecular Films. A maximum value of the size of atoms has been obtained by using the following considerations. The thinnest layers of a substance, which still retain the properties of that substance, have at least a thickness which equals the diameter of an atom; but it is more likely that these films will be several atoms thick. **Faraday** had already prepared the thinnest of gold leaf which is used for gilding various objects. By weighing the whole leaf and determining its area, he was able to show that the thickness was less than one hundred-thousandths of a millimetre. Later, experiments by more refined methods led still further. A known mass of oil was allowed to spread over the surface of water and the thickness of the film was measured by determining its area as with the gold leaf; in 1912 **Devaux** found that the thickness of these layers varied from one to one and a half millionths of a millimetre. Several other investigations by **Langmuir** and **N. K. Adam** have confirmed these values. We can therefore say with certainty that the maximum diameter of a molecule is less than 10^{-6} mm., probably very much smaller.

Molecular Velocities in Gases. Let us for a moment consider those properties of gases, which for reasons of the proposed hypotheses we have ascribed to molecules, and which offer the earliest prospects of being detected experimentally. Without doubt they will be those which take part least in the extreme smallness of the particles relative to dimensions which are familiar to us. This will be fulfilled by the molecular velocity, which for some gases attains to high values.

If a large vessel is filled with gas and its wall pierced at any point by a fine hole leading to an empty chamber, the gas molecules which fall upon the opening will leave the vessel; if care is taken to maintain the original pressure inside the apparatus in spite of this leakage there can be no doubt that the number of molecules which pass through the hole is equal to the number of molecules which would otherwise have collided with this part of the wall, and that the velocity of the stream is equal to the velocity of the gas molecules. The number of impacts per second upon unit area is $N/3 \times v/2a^3 = Nv/6a^3$, if we again assume the space to be cubical [*cf.* p. 119 *et seq.*]. If we further assume that a second gas fills the vessel under the same pressure, according to Avogadro's law it will contain N molecules, so that the number of impacts per unit area per second is $Nv'/6a^3$, where v' is the molecular velocity for this second gas. This implies that the number of molecules passing through the opening is proportional to their velocity, and since this is inversely proportional to the square-root of the density of the gas, it follows that the number of molecules issuing from the hole is proportional to the square root of the density of the gas. The streaming of a gas through a hole is quite similar to the process of diffusion, it really being the limiting case in which the free paths have become infinite. These assertions concerning the quantity of gas which passes through the orifice have been confirmed experimentally.

We now wish to find, however, some process by means of which the motion of the molecules can be traced by our eye. The escaping molecules must, of necessity, leave the orifice with a velocity with which they would otherwise have struck the wall of the vessel, *i.e.* their velocity is v. In general we cannot measure this velocity, since the molecules collide with air molecules and are deflected from their path. These latter collisions can be prevented by allowing the gas to escape into a vacuum. Under these conditions the molecules will retain their velocities, and we might even conceive an experiment to measure these velocities, so that not only could molecular motion become an established fact, but the calculated velocities would become subject to direct confirmation. Such a proof would be very important, for it would then mean that things which had previously only been discussed theoretically could be counted amongst those things which have been observed.

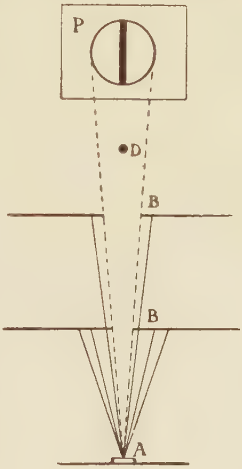


Fig. 87.

These experiments were first carried out by **Dunoyer** in 1911. He allowed a piece of metallic sodium A, Fig. 87, to evaporate into a vacuum. This vapour passed through two narrow slits BB placed one behind the other, afterwards being condensed upon a glass plate P. If theory is correct, that the molecules move with a great velocity along rectilinear paths until they collide with other particles, then it is only those particles moving in the direction AB which will reach the glass wall. Such molecular impacts are here prevented by performing the experiment in a very high vacuum. Finally, the particles will meet the glass and form a metallic mirror there. This mirror will have exactly the same shape as the image formed when light from a source at A is allowed to pass through the slits. If a thin wire is placed at D so that it intercepts some of the sodium molecules, its presence will be indicated on the glass wall by the presence of a 'shadow', *i.e.* a region free from condensed sodium. These expectations were completely confirmed by these experiments which formed the starting point for a direct experimental treatment of the atomic theory. In Fig. 87 the glass plate P is shown in such a position so that the sodium mirror, and the 'shadow' of the wire, can be seen readily.

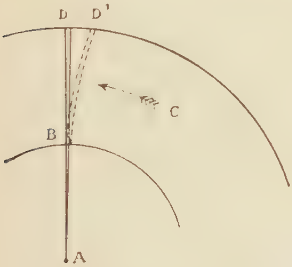


Fig. 88.

In 1920 **Stern**, using an apparatus very similar to that just described, measured the velocity of gaseous molecules. The main points will be appreciated from Fig. 88. Let A be the source of a stream of atoms; *i.e.* A is the centre of a thin silver wire heated by means of an electric current so that it evaporates slowly, just as the filament of an ordinary electric bulb (of the vacuum type) evaporates and blackens the bulb when the current is too large. Let B be a narrow slit in a double-walled metallic cylinder C capable of rotation about an axis passing through A. The silver atoms strike a cylinder made of polished copper at a point D, where

ABD is a straight line. If now the cylinder C with the slit B rotates as a whole, then in the time occupied by the molecules in traversing the distance BD the point D' on the copper cylinder will have arrived at D, and the silver mirror will appear not at D, but at D'. This allows the velocity of the molecules to be calculated, for they will have crossed the distance BD in the time taken by the cylinder to move from D to D'. If v is the velocity of the molecules, and c that of the copper cylinder, then

$$v : c = BD : D'D.$$

The cylinder was driven by an electric motor making n revolutions per second. If r is the radius of the outer wall of the cylinder C, then the velocity of any point upon it is $c = 2\pi rn$. The distances BD and $D'D$, being known, we have

$$v = 2\pi rnBD/D'D.$$

According to the kinetic theory the velocity of the silver molecules at the temperatures used in these experiments is 672 m/sec. In one series Stern found 643 and in another 675 m/sec. The idea that the pressure of a gas was due to the kinetic energy of its molecules was confirmed completely, for the velocities of the molecules in a silver vapour had been measured and found to agree with those calculated from theory.

The Number and Paths of Single Atoms. Several radioactive phenomena have shown to us directly that matter possesses an atomic structure. With an instrument known as the spinthariscope, the individual atoms emitted from radium during its disintegration fall upon a screen coated with zinc sulphide. Each impact produces a bright flash of light which can be seen with no other aid than that of a low power magnifying glass by an eye accustomed to the darkness.

Rutherford and Geiger succeeded in increasing the ionizing effect of a disintegrating atom of radium to such an extent that the total charge of all the ions could deflect an electrometer needle. The arrival of each new atom was observed, so that they could be counted individually, or shown to an audience. In this way the number of atoms in a gram of radium has been calculated.

Finally, **Wilson** has succeeded in making visible to us the path traversed by an alpha-particle when it is ejected from radium. The alpha-particle has been shown to be an atom of helium carrying a positive charge. The ions which were produced when the alpha-particle passed through a saturated water vapour formed the nuclei upon which condensation of the vapour could occur. The phenomenon was recorded by means of instantaneous photographs which have yielded much information concerning the collisions between alpha-particles and air molecules. The author has carried out these last two experiments with simple apparatus so that they can now be shown in the course of the usual school curriculum. They will be described in the chapter on radioactivity.

2. THE DETERMINATION OF ATOMIC CONSTANTS

Having obtained the information detailed above, the atomic theory ceases to be a hypothesis. Its assertions are now facts, which can be subjected to experimental confirmation at any time. This, however, has given rise to many more new questions, the answers to which would permit us to gain a deeper insight into the secrets of the material world. Since, so to speak, we can see the atoms themselves, is it not possible to discern experimentally the sizes of the different atoms, and to express their mass in grams, etc.?

Such a wish was fulfilled almost at once, for, with relatively little difficulty, we were able to calculate at once the velocity with which the molecules in a gas moved at different temperatures merely by assuming that the pressure of the gas was due to molecular collisions.

If, however, the particles are all moving with unknown velocities until they collide with other particles, then how great is the distance traversed during successive encounters, how often does a particle change its direction every second and how often does it collide with others? Let us consider for a moment how deeply these questions fathom the internal mechanism of matter; these questions may be compared with the discussions which the ancient Greeks held concerning a few very superficial phenomena. To-day, we ought really not to ask these questions, for they have been solved with more or less certainty, so that the validity of these solutions can no longer be doubted.

Moreover, in approaching the reality, there are always two degrees of approximation to be distinguished. In the first stage we make the simple assumption that the velocity, free-path, and number of collisions, are the same for all the molecules; we try to find *mean values* for all these quantities as we have already done in the case of their molecular velocities, *i.e.* we seek to discover the velocity which every molecule must have in order to exert the same pressure on the walls of the containing vessel as the molecules with their unequal velocities actually do. Similarly the mean free-path is that mean distance through which all particles would have to move during successive encounters in order to account for all the phenomena quantitatively; actually the free-paths are not equal but, in an unknown manner, are more or less equal to this mean value.

In the second approximation to the reality we ask how we have to conceive the distribution of the real velocities, free-paths and collisions around these mean values. That does not mean that we can speak about the atoms individually, for that would assume that an individual atom could be recognized again; a thing which is obviously beyond our powers. Nevertheless, we can still seek to find what portion of the molecules will have velocities (or etc.) differing by a certain amount from the mean free-path.

The Mean Free-path. Different methods are now known by means of which atomic constants can be ascertained; the oldest of these is one by which the mean free-path of molecules in a gas was first determined. Since this determination of the mean free-path has an inherent value, the method which was used for this purpose will now be discussed briefly. The mean free-path is calculated from observations made on the coefficient of viscosity of a gas, so that this property of a substance must now be considered.

Let us assume that we have a liquid on the surface of which is situated a plate which the liquid wets. If the plate moves along the surface of the liquid with constant velocity, then a definite force is necessary to overcome the resistance which is offered to the motion.

This resistance is not due to friction between the plate and the liquid, but to friction between the different layers in the liquid. Exact experiments, in which the plates were wetted by the liquid, have shown that the plate carries with it the layers of liquid which lie next to it. This fact can very easily be verified. There is therefore no relative motion taking place between the plate and the layer of liquid immediately adjacent to it; the motion takes place between this layer of liquid and the layers which are situated at greater depths in the liquid. In these experiments we therefore measure the force necessary to produce this relative displacement amongst the various layers in the liquid. This force is due to the **viscosity** of the liquid. Now in gases there are said to be no forces between the individual molecules so that one might expect that gases would not show any viscous properties. But gases do exhibit these characteristics. Suppose that a light plate, A, Fig. 89, is mounted on a pivot about which it can rotate; at a distance of about 1 cm. above it a second small plate, B, is rotated. In spite of the distance between these two plates, after a short time, the lower one will be found to be rotating. If the experiment is made in vacuo the lower disc will not rotate. The phenomenon therefore depends upon the nature of the gas in the space between the two plates. When the experiment is repeated with some other gas, then the magnitude of the rotation produced is different. It is therefore caused by the moving gas molecules putting neighbouring molecules into motion, so that we have to assume that gases are characterized by a coefficient of viscosity. We must now give some account of the way in which viscous forces arise in gases.

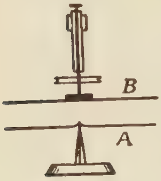


Fig. 89.

The molecules of a gas are generally moving irregularly in all directions, no special direction having any preference over the others. When the gas as a whole moves in one direction, then this ordered motion is superimposed on the irregular motion which already exists. If now a plate moves with definite velocity in a gas, the adjacent layers of gas partake of this motion. The more distant layers move with a smaller velocity. On account of the perpetual thermal agitation of the molecules there will always be molecules coming from the lower regions in the gas and moving upwards towards the plate. They arrive with a velocity which is less than that of the plate, so that they must continually be acquiring a greater velocity in virtue of the plate. On the other hand, molecules having a high velocity leave the region near the plate, and pass into the lower layers of the gas, which gradually tend to move with a velocity equal to that of the plate. The molecules which pass from near the plate have a greater velocity than those which arrive to take their place, so that the plate loses energy continuously. If, in spite of this, the plate is to move with constant velocity, then this loss of energy must be supplied from without. A force is therefore necessary to maintain the plate in motion.

Instead of giving a formal proof we shall only indicate upon what circumstances the magnitude of the force depends, which drives the plate. The coefficient of viscosity is defined as the force acting upon a plate whose area is 1 cm.^{2} when the relative velocity between it and the layer of gas 1 cm. distant from it is 1 cm. per sec. From the explanation given above of the way in which viscous forces arise three conclusions may be made: (a) the viscosity will be greater, the greater the mass of the molecules set in motion by the disc, *i.e.* the viscosity is proportional to the density; (b) the faster the molecules move, the greater will be the number accelerated per second, for the greater will be the number of molecules which approach and recede from the plate per second; (c) and this is for us the most important point, the difference between the velocity of the approaching molecules and those of the plate will be greater, the greater the distance from which the molecules reach the plate without colliding on their way, *i.e.* the greater is the mean free-path of the molecules of the gas. A rigid calculation has shown that

$$\eta = \tfrac{1}{3}dvL \quad . \quad . \quad . \quad . \quad . \quad (87)$$

where η is the coefficient of viscosity, v the linear velocity of the molecules, L the free-path, and d the density of the gas.

In order to measure the viscosity of a gas, a round disc suspended by means of a quartz fibre, so that it can oscillate in a horizontal plane, is used. The apparatus is similar to the torsion pendulum shown in Fig. 22, p. 54. The oscillations die down much more quickly when the disc moves in a narrow space containing gas, than when it oscillates in vacuo. The difference is due to the viscosity of the gas so that the variation can be used in order to measure the coefficient of viscosity. We shall therefore assume that η is known. The velocity of the molecules has been found [*cf.* Table 23, column 3] and the density of the gas is also known. The only unknown in equation (87) is therefore L, the value of which can now be calculated. The same table gives in column 4 the mean free-path for several gases at 0° *C.* and 76 mm. of mercury pressure. **Maxwell** concluded from the kinetic theory of gases that the viscosity is independent of the pressure of the gas, a fact which has been confirmed over a wide range of pressures. This agreement between theory and experiment has been urged as a strong support for the validity of the kinetic theory of gases. Equation (87) indicates it too, for at lower pressures the density becomes less and the free-path correspondingly longer. The above conclusion cannot be absolutely true, for at very low pressures the viscosity must become zero, since there will be no molecule left.

TABLE 23

The Molecular Constants of Some Gases

Gas.	Formula.	v in m/sec.	L in mm. $\times 10^6$.	z per sec. $\times 10^{-6}$.	σ_0 in mm. $\times 10^6$.
Hydrogen	H_2	1,692	112	15,100	0·23
Helium	He	1,204	180	6,600	0·19
Nitrogen	N_2	454·3	60	7,570	0·31
Oxygen	O_2	425	64·7	6,570	0·29
Methane	CH_4	600	49·3	12,160	—
Water vapour	H_2O	566	40·4	14,000	0·26
Argon	A	381	63·5	6,010	0·28
Neon	Ne	535·1	125·8	4,260	0·23
Hydrogen-chloride	HCl	398	43·3	9,200	—
Carbon monoxide	CO	454·5	58·4	7,780	0·32
Carbon di-oxide	CO_2	362·5	39·7	9,150	0·32
Chlorine	Cl_2	285·6	28·7	9,950	0·36
Krypton	Kr	264·1	48·7	5,430	0·32
Xenon	X	211	35·0	6,230	0·35
Ethylene	C_2H_4	454·2	34·5	13,150	0·36

v = mean velocity in metres per second at 0° C.
L = mean free-path at 0° C. and standard atmospheric pressure.
z = number of collisions for one molecule per second $= \frac{v}{L}$.
σ_0 = diameter of a molecule.

The Number of Molecular Collisions per Second. From a knowledge of the mean free-path it is possible to calculate the number of collisions per second in which a gas molecule takes part. We already know the velocity v of the molecules, *i.e.* we know the distance traversed in one second, and now that we have learnt the distance L through which a molecule moves between successive encounters, the number, z, of collisions per second is known for $z = v/L$. The value of this number for several years is shown in the fifth column of the above table.

At a late date (1920) **Born** succeeded in measuring the values of the mean free-paths by using some ingenious and carefully designed apparatus. The experiments are very similar to those in which the molecular velocities were determined directly. Born first determined the quantity of silver which was deposited per second upon a small plate when silver was heated in vacuo to a definite temperature. The quantity so deposited was determined for several different distances of the plate from the source of atoms. Air was then allowed to enter, but all other conditions kept constant. The quantity of silver deposited for the various plate distances was now diminished, for every silver atom which collided with an air molecule in its path from the source of atoms was deflected from its course. In this way he was able to calculate the mean free-path and the values so obtained were consistent with those calculated by theoretical considerations. The mean free-path must henceforth be counted amongst those facts which can be observed.

The Number of Molecules in a Gram-molecule. We are now in a position to calculate one of the most important of physical constants. For the time being this achievement marks the crowning point in our theory of the atom, for we are able to assert how many atoms are present in a given mass of substance. From the beginning this has been the goal to which the eyes of physicists and chemists have always been directed. The very oldest attempts to find the diameter of a molecule by measuring the thickness of gold-leaf, or layers of oil upon water, enabled us to say that this number was at *least so big*. It gave us a lower limit to this number. The first determination, which was at least of the right order of magnitude, was made in 1865 by **Loschmidt**, a physicist in Vienna. He knew that the number of molecules per cm.^{3} could be ascertained from a knowledge of the mean free-path. This number is referred to as **Loschmidt's Number.** Since we know that all gram-molecules of gases occupy 22·4 litres at normal pressure and temperature, we will determine the number of molecules in 22·4 litres, for this number is of universal importance. It is valid for all gases and, in general, for the gram-molecules of all substances.

In order for a molecule to move freely over a distance L, it is necessary that it should not collide with any other molecule in this distance. If this distance is represented by a straight line over which the centre of gravity of the molecule moves, then the centres of gravity of neighbouring molecules must be at least $2r = d$ cm. from this line, where r is the radius of the molecule. The whole space which is available for each molecule is at least therefore equal to that of a cylinder of length L and cross section πd^2, *i.e.* the volume is $\pi d^2 L$. But if the molecules are moving, the collisions will take place more frequently. Taking this into account, **Maxwell** showed that this volume must be increased $\sqrt{2}$ times, so that the space in which each molecule is free to move is $\pi d^2 L \sqrt{2}$. If V is the total volume of the entire N molecules, then $N\pi d^2 L\sqrt{2} = V$. This equation still contains two unknowns, N and d.

In order to discover a second equation having the same unknowns in it, Loschmidt made use of the volume which a gas fills when it has been liquefied. At most the molecules cannot more than touch one another when in this condition. If the molecules are considered to be cubes the length of whose edge is d each molecule will fill a space d^3, and if these molecules fill the whole of the space V_L, occupied by the liquid, then $Nd^3 = V_L$. If the molecules have some other shape than that now assumed, *e.g.* if they are spherical, then the space will be occupied differently, but the order of magnitude of the number we are seeking will not be affected by it. In this way Loschmidt obtained a second equation between these unknowns so that N and d could be calculated. But the number was still inexact, since the molecules of a liquid do not fill the space occupied by the liquid completely, for otherwise such motions as those we see in the Brownian movement would be impossible. Proceeding in this manner he found, for 1 cm.^{3}, $N = 1{\cdot}31 \times 10^{15}$, so that for 22,400 cm.^{3} the number is 29×10^{21}. For carbon di-oxide the diameter is $16{\cdot}2 \times 10^{-8}$ cm. The first number is about twenty times too small, the second is five times too large, but they are of the right order of magnitude.

The Determination of Loschmidt's Number by means of Brownian Motion. The connecting link between the Brownian movement and the whole series of atomic constants was made in two stages. **Van't Hoff** first showed that the gas laws were valid for dilute aqueous solutions, *e.g.* the sugar molecules dissolved in water exert the same pressure in the water as they would do if they were present as a gas occupying the same volume as the solution. **Einstein** extended these ideas so that they applied to small particles suspended in an emulsion. Like the molecules of a gas they collide with one another; they are set in motion, even participating in a Brownian movement, and, on account of their motion, they exert a pressure on the walls of the containing vessel. Einstein gave the whole theory of this phenomenon and succeeded in discovering several laws by applying the theory of probability to the completely irregular motions which occur. In Paris **Perrin** has made experiments on these lines and was not only able to reveal the motion of the suspended particles, but was also able to ascertain the number of molecules in a gas. For this purpose equation (55) was used. The solid particles which are in suspension will tend to collect at the bottom of the liquid, and the number of particles at different heights in the liquid will be given by this same equation. The change will take place more rapidly in this case than it does in the air on account of the greater mass of the particles. The equations (54) and (55) which we now use are changed a little in order to suit the present problem. The density d_0 of equation (55) denotes the mass per unit volume at pressure p_0. This mass can be replaced by the expression mN/V_0, where m is the mass of a particle, N the number of such particles occupying a volume V_0. If V_0 is the volume of a gram-molecule, viz. 22,400 cm.^{3}, then N is the number of particles in a gram-molecule. In doing this we have, with Einstein, tacitly assumed that, amongst others, the law of Avogadro is true for the particles suspended in an emulsion, *i.e.* the number of molecules present is the same as the number of molecules occurring in a volume of gas equal to that of the liquid under the same pressure and at the same temperature. We therefore have to write Nm/V_0 instead of d_0, so that d_0/p_0 becomes Nm/V_0p_0.

But according to the laws of gases, when we consider a gram-molecule, $p_0V_0 = RT$, where R is a universal constant and T is the temperature measured on the absolute scale [*cf.* equation (75), p. 149]. The equation (55) thus becomes

$$\frac{z_H}{z_0} = e^{-\frac{NmgH}{RT}}$$

or, taking logarithms to the base e, we have

$$\log_e \frac{z_H}{z_0} = -\frac{Nmg}{RT}H \qquad . \quad . \quad . \quad . \quad (88)$$

Using this equation **Perrin** has determined Loschmidt's number. Since it is only necessary to know the ratio z_H/z_0, it is sufficient to determine the number of particles first at any arbitrary position and then at a distance H above. Perrin observed the particles by means of a microscope the field of which was limited with a suitable diaphragm. The microscope was placed in a vertical position, the distance H therefore being equal to the distance through which the microscope was raised. Perrin made a very large number of experiments in order to obtain the best possible mean value. For example, he found, as the result of 13,000 measurements at four different heights, all at a distance 0·03 mm. above one another, the numbers 100, 47, 22·6, 12. According to the law which states that the ratio of the number of molecules present at any two successive points differing in height by the same amount is constant, he should have obtained the numbers 100, 48, 23, and 11·1. Considering the difficulties which attend the measurement of such small distances as are here involved, the agreement between these two sets of numbers is excellent. The agreement also affords a splendid confirmation of the fundamental conception made by Einstein.

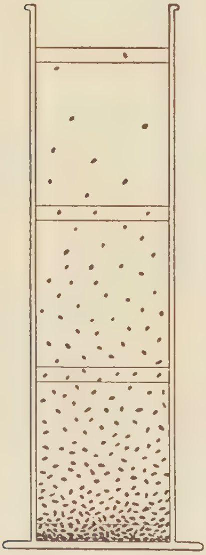


Fig. 90.

The only factor in the above equation which is still unknown is m, the mass of a suspended particle. Perrin procured for himself particles which were as nearly equal to one another as possible, and then determined their mass in several different ways. For his experiments he used small grains of *gamboge* or *gum mastic*. Since he knew the density of these substances it was only necessary for him to determine the diameters of the suspended grains; this was made possible by using a microscope. In the experiments which were cited above the grains had a diameter of 0·212 μ, so that the mass m was 48×10^{-15} grm. Now $R = 8{\cdot}3 \times 10^7$ ergs., according to the kinetic theory of gases, and T the prevailing temperature of 20° C. is 293° K.

The diameter of the grains was determined by measuring their rate of fall in a liquid. They will not move with an acceleration equal to that due to gravity acting on bodies in air (strictly in vacuo), *i.e.* the force on each particle is not mg owing to the upthrust of the liquid which the particle displaces. If v is the volume of a single grain, then if D is its density, its mass will be Dv, so that the force of attraction due to the earth upon it is Dvg. The mass of the displaced liquid is dv, where d is the density of the liquid, so that the upthrust on the grain is dvg; this will make the apparent weight of the grain less, so that the resultant force acting downwards on the particle is $f = gv(D - d)$. Using equation (8) to calculate the acceleration, a, of the grain in the liquid, we have

$$a = \frac{f}{m} = \frac{gv(D - d)}{Dv} = g\frac{D - d}{D}.$$

This is the acceleration which must be used in equation (88) instead of g.

In the experiment we have already referred to, Perrin found $D = 1{\cdot}194$, and since the solution was an aqueous one, $d = 1{\cdot}00$. For $\frac{z_H}{z_0}$ he found from the above figure the result $1/2{\cdot}03$. Now $\log_e 2{\cdot}03 = 0{\cdot}708$. Inserting these values in equation (88) we have $N = 70 \times 10^{22}$ approximately.

These experiments have been repeated by other observers using the same and different pieces of apparatus. Counting only those methods which are quite free from objection, there are about twenty-five in all. The results of these different methods differ as little as is to be expected from the difficulty of the observations. As the final result of the best of all the measurements we have the value $N = 60{\cdot}5 \times 10^{22}$. The true value of N does not appear to differ from this by more than one per cent. The various determinations have involved different branches of physics, *e.g.* the Brownian movement, the kinetic theory of gases, the scattering of light by gaseous molecules, the radiation from a black body, and the determination of the smallest particle of electricity, combined with the results of electrolysis. Finally radioactive phenomena have supplied the means whereby several other experiments could be made. These are independent of any complicated theory and therefore easy to interpret; they will shortly be described. Even if several methods contain many assumptions which have not been proved, the fact that they furnish consistent results is sufficient evidence for us to regard them as being reliable.

Moreover, the last-mentioned processes, which occur in connexion with radioactive substances, are so direct and free from all assumptions, that one could say they are really processes in which the atoms are actually counted. But the number of atoms in a gram of substance is so infinitely big compared with the age of an average man, that one has to be satisfied by counting the number in a known small mass, the desired number being obtained by multiplication. In order to conceive the enormous number of molecules in a gram-molecule, let us assume that we have a machine capable of counting 200 million molecules in a second. The year contains about 30×10^6 seconds, so that the number of molecules which could be counted in a year is $200 \times 30 \times 10^{12} = 6 \times 10^{15}$; hence 100 million years would elapse before all the molecules in 2 grm. of hydrogen could be counted.

The Diameters of Molecules. At this point we find ourselves in a position ready to discuss the remaining question: How big is an atom of hydrogen or oxygen? There are about fifteen different methods, each independent of one another, by means of which the size of an atom can be estimated. We shall only refer to one which has already been mentioned. On p. 125, whilst discussing the deviations from Boyle's law as represented in the equation of Van der Waals, the magnitude of b was found. It is equal to four times the actual volume of the gas molecules. Since we now know the number of molecules in any volume, the volume of one molecule is easily computed by division, and, assuming that the molecules are spheres, their diameter is then obtained. For so long as we remain ignorant of the actual shape of an atom, by introducing the idea that it is a sphere, we assign a shape which involves the least number of particular and unproved assumptions. The last column of the table on p. 207 contains the calculated values. The diameters are of an astonishingly small order of magnitude, being only a few ten-millionths of 1 mm. This magnitude is of course only estimated, for there is no microscope which can make such a small length visible to us. The diameter is too small to be appreciated, for we have no apparatus capable of measuring such a length. Fortunately, however, this conclusion, which is derived from measurements involving the volume of a large mass of gas and the number of molecules per unit volume (which is now known so well), is so simple and convincing that we are no longer able to doubt the validity of it.

The Mass of an Atom. Having learnt the number of atoms in a gram-atom, *e.g.* in one gram of hydrogen, it is easy to calculate the mass of an atom of hydrogen. Since $60 \cdot 10^{22}$ atoms of hydrogen have a mass of 1 grm., it follows that 1 atom has a mass of $1 \div 60 \times 10^{22} = 1 \cdot 66 \times 10^{-24}$ grm. ; and an atom, m_A, of atomic weight A, has a mass

$$m_A = A \times 1 \cdot 66 \times 10^{-24} \text{ grm.} \quad . \quad . \quad . \quad (90)$$

The Kinetic Energy of a Molecule. A very important result concerning the energy of molecules may here be noted. According to equations (56) and (75),

$$PV = RT = Nmv^2/3.$$

The universal gas constant R is proportional to N, the number of molecules present, so that R can be calculated and interpreted for each molecule. Denoting its value by k, we have

$$k = \frac{R}{N} = \frac{8 \cdot 3 \times 10^7}{60 \cdot 5 \times 10^{22}} = 1 \cdot 37 \times 10^{-16} \text{ ergs per degree} \quad . \quad (91)$$

This equation shows us at once that k is equal to two-thirds the kinetic energy of a molecule at a temperature of 1°. This constant k is known as **Boltzmann's** constant. The energy due to its rectilinear motion amounts in the case of a single molecule to

$$E_k = \tfrac{3}{2}kT = 2 \cdot 06 \times 10^{-16}\ T \text{ ergs} \quad . \quad . \quad . \quad (92)$$

when the temperature is $T°$ absolute. This value is quite independent of the type of molecule which has been considered, for the energy due to rectilinear motion is equal for all gases at the same temperature and having the same number of molecules.

3. The Distribution of Molecular Velocities

Finally we have to consider the question, how are the velocities of the individual molecules related to their mean velocity ? Have all the molecules velocities which are very nearly equal to this mean value, or can some of them have velocities which are very different from it ?

The fact that the molecules must have different velocities is very easily seen. Even if we were to think that at some definite instant all the molecules had the same velocity, shortly afterwards they would again be very different. For amongst the $60.10^{22}.10^{11} = 6.10^{33}$ molecular collisions which occur in a gram-molecule of hydrogen per second, it may be assumed that one molecule is hit by two or three molecules travelling in approximately the same direction at the same instant. It will then behave like an elastic sphere with which a heavier sphere has collided, *i.e.* its velocity will be increased considerably. Or we may assume that only two particles, K and K', Fig. 91, both moving with velocity v in perpendicular directions, collide, but in such a way that one is struck in its direction of motion AB, whereas for the other, K', the direction of collision AB is perpendicular to its direction of motion CD. After the collision the sphere K will have been brought just to rest, whilst the sphere K' has a velocity the rectangular components of which are each v, so that its resultant velocity is $v\sqrt{2}$.

Moreover, such a proof is no longer necessary, for in the experiments by **Stern,** already described on p. 203, the velocities of the molecules were actually found to be different. Stern observed that a sharp image of the slit (formed of silver atoms) was obtained when his cylinder was stationary. When it was rotated, however, the image became somewhat indistinct along its two sides, thereby proving that some of the molecules were travelling with different velocities. By making use of certain assumptions it became possible for the distribution of molecular velocities around their mean value to be calculated. The number travelling at any one particular speed is proportional to the mass of silver deposited at the corresponding position in the image. If this could be done, the problem would have been subjected to direct experimental confirmation. At present, however, the necessary technique has not been acquired, for the experimental difficulties are very great.

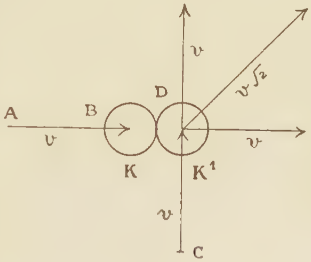


Fig. 91.

Maxwell solved this problem mathematically by applying to it the theory of probability. He assumed that the molecules were in a state of complete chaos, *i.e.* if the velocities of all the molecules are resolved into three components at right angles to one another, then these components always have values lying between 0 and v. (It is zero when the direction of its motion is perpendicular to that of the axis of resolution, and v when its motion is along this direction.) These three components are denoted by v_x, v_y, v_z. In a motion which is absolutely chaotic.

The velocity of every particle is always given by

$$v = \sqrt{v_x^2 + v_y^2 + v_z^2}.$$

Maxwell then assumed, that when the number of molecules having a definite velocity v is very great, the more probable it becomes that this particular velocity will exist in the gas. He succeeded in discovering the law which governs the probability of such an event, and his methods were so remarkable that it was not necessary for him to consider the laws of impact or of gases. His solution is therefore valid under all similar conditions where there is a large number of values distributed irregularly round a mean value. Maxwell's formula was therefore exactly like the one which **Gauss** had obtained, when calculating the probable error from a series of experimental values distributed around the true value.

Maxwell also calculated the probability of a molecule being found whose velocity should lie within the limits v and $v + dv$, where dv represents a small increase in the velocity. He calculated the probability for all values of v lying between 0 and ∞. He showed that there was a velocity, which we will call a, for which the probability was a maximum, *i.e.* it was most likely to occur. The velocities which are greater or less than a occur less frequently ; in fact, they become more rare the greater the difference between them and a. For $v = 0$ and $v = \infty$ the probability is zero. Expressed somewhat more exactly we may say that for velocities which only differ by an amount dv from these extreme limits, the probability is vanishingly small. Maxwell's formula is

$$W = \frac{4}{\sqrt{\pi}} \frac{v^2}{a^2} e^{-\frac{v^2}{a^2}} \frac{dv}{a} \quad . \quad . \quad . \quad . \quad (93)$$

where e is the base of natural logarithms. If we write $v/a = z$, so that a, the velocity for which the probability is a maximum, becomes our unit of velocity, the formula reduces to

$$W = \frac{4}{\sqrt{\pi}} z^2 e^{-z^2} dz \quad . \quad . \quad . \quad . \quad (94)$$

The meaning of this result is obtained at once if it is represented graphically. The horizontal axis, Fig. 92, is chosen as the axis of z, and at each point along it the corresponding value of W is plotted, where W is the probability that a certain value will occur. The probability is a maximum when $v = a$ and is then equal to 0·8. If dv is made so small that the curve within this region does not deviate by any appreciable amount from the maximum value 0·8, then $0{\cdot}8 \times \frac{dv}{a}$ will represent the fraction of all the molecules having velocities which lie between a and $a + dv$. Each element of area on a base dz and of height z represents the number of molecules whose velocities lie between z and $z + dz$. It is easily seen that the number of molecules, for which $v = a$ approximately, forms a very large fraction of all the molecules present. The mean velocity is therefore very nearly equal to the velocity of the greater portion of the molecules present.

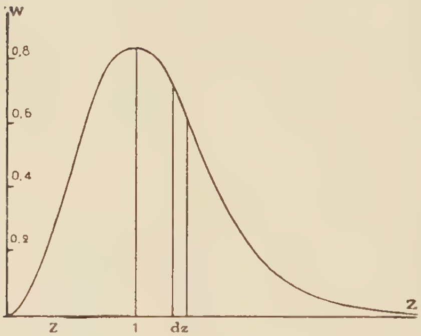


Fig. 92.

The following table contains the number of molecules with different velocities when 1000 molecules of oxygen are present in a volume at 0° C. The mean velocity in such a case is $v = 425$ m/sec.

Table 24

The Distribution of Molecular Velocities

Velocity (m/sec.).	No. of Molecules.
0–100	32
100–200	82
200–300	167
300–400	215
400–500	203
500–600	152
600–700	92
over 700	76

The true free-paths will also be grouped round the above stated mean free-paths according to the same law.

4. ISOTOPES

The results of the discussions we have so far had may be summed up as follows : in many cases we have succeeded in measuring molecular velocities, mean free-paths and the number of molecular collisions per second, but our results have always been in the form of mean values, around which the actual values for individual molecules are distributed very closely. At the same time we have learnt the proportion of molecules present having a definite deviation from this mean value, so that our desire for an insight into the internal structure of matter has been wellnigh fulfilled. Only in one instance is some further information desirable, and that concerns the atom itself. Are the numbers which we have obtained for the weight of an atom also mean values which are more or less equal to those of the individual atoms, or has every atom of hydrogen the same weight ? From the very beginning this question is more fundamental than that which concerns the velocity. For the velocity of a molecule is something which is quite fortuitous and can be changed very easily when a molecular encounter occurs. In the same way the mean free-path depends upon the relative positions of the molecules and consequently upon the collision which a neighbouring particle has made during the previous instant. This is not so, however, in respect to their mass. We know that the mass of a body is independent of its physical condition, and that it can only be changed by dividing the body up into smaller fragments. Our conception of an atom itself would be essentially different if we were to regard it as a mixture of particles having a definite mean mass rather than as particles of exactly the same mass.

In 1803 **Dalton** postulated that the atoms of one and the same element are indistinguishable from one another and equal in mass. We must now remember how, in those days, a decision had to be made as to whether an unknown substance were an element or not. When it was not possible to divide a substance into two components each exhibiting different chemical and physical characteristics, that substance was called an element. In other words, it was only the assertion that atoms of one and the same element were identical that enabled these early workers to decide whether or not they were dealing with an element. The remaining part of Dalton's postulate, viz. that the atoms were equal in mass, was purely hypothetical, but nevertheless a very natural assumption. For the idea that substances, whose atoms were different in mass, should exhibit the same physical and chemical properties was very improbable : after the lapse of a few decades, during which many new investigations had been carried out, it was still less probable than at the time of Dalton. Nevertheless, a hundred years after Dalton research workers were compelled by discoveries made in connexion with radioactive substances to consider again the question as to whether all substances, which, according to the views of a chemist, were elements, and which had the same atomic weight by whatever chemical reaction it was measured, were not really mixtures of particles having different masses.

The reasons which caused this problem to be raised will be dealt with later on when discussing radioactive phenomena ; for the present we shall simply have to accept the facts. In particular, the suspicion had gradually been growing that lead was not an elementary substance and that its atomic weight would depend upon the substances in whose vicinity it was found in Nature. A redetermination of the atomic weight of lead became very necessary, especial attention being paid to the source from whence it was obtained—a factor which had previously been disregarded. When the experiments were conducted with the utmost care the following results were finally obtained in 1914 and the following years by **Hönigschmid** in Vienna. The atomic weight of lead obtained from the mineral thorite (Norwegian) was 207·9, whilst that obtained from pitchblende (from Morogoro, Tanganyika Territory) gave 206·046 ; the usual form of lead, which up till then had been the only type investigated, had an atomic weight 207·2. It is very certain that this result had a most important influence on our ideas of atoms, but at the same time it increased rather than answered the problems of that day.*

The question as to whether the atomic weights were mean values had not yet been solved. Did the two masses of lead which had been investigated contain in the one instance an excess of larger atoms, and in the other of smaller ones ? For if so the earlier value would still be the mean value around which the weights of the individual atoms were distributed. Or are there two kinds of lead, one having an atomic weight 206 and the other 208 ? If there are two sorts, the same question can again be asked : Are the numbers 206 and 208 mean values, each for one sort of lead, or are there two kinds of lead atoms, one with a fixed atomic weight 206 and the other 208 ?

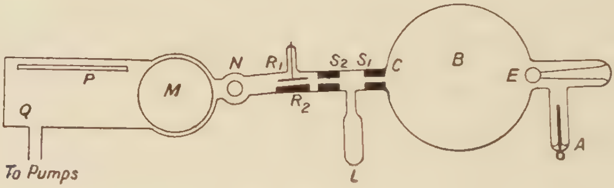


FIG. 93.

Aston's Mass Spectrograph. These critical problems were answered by a series of researches made by **Aston** in 1919 and the following years. The essential features of his apparatus are shown in Fig. 93. B is a large glass bulb through which an electric discharge passes between the anode A and the concave cathode C. The cathode rays which are thereby produced are focussed on to a quartz bulb E, so that a fusion of the glass walls of B is prevented from taking place. The positive rays from B pass through a narrow slit S_1 in C, whilst the second slit S_2 serves to define a narrow beam of the rays. The space between these two slits is highly exhausted by means of charcoal contained in a bulb L kept at the temperature of liquid air. The narrow beam of positive rays from S_2 then passes between two plates R_1 and R_2 between which an electric field exists. This causes the beam to be spread out into a broad band, part of which is allowed to pass through an adjustable slit N. The rays then pass between the poles of a large electromagnet M which is so arranged that they are deflected in an opposite direction to that produced by the electric field. They then

* Actually, at that time, many new discoveries were made which helped to throw some light on this question. In order not to burden with too many details the reader who is becoming acquainted with these facts for the first time, only points of historical importance are given in the above account. In view of the statements which have already been made, a careful reader should find no difficulty in following the argument.

pass into a chamber Q, where they fall upon a photographic plate P. This arrangement causes all particles having the same mass to strike the plate at one definite point. **Aston always obtained at least one sharp image of the slit,** but very often there were several well-defined images. From these observations the following very important conclusion is made: **the atomic weights are not mean values in the sense that the different particles have masses more or less approaching that of the mean value, but there exist, in some elements, atoms of a different kind.** Each set of atoms has a definite atomic weight which is the same for each atom of that kind but different from the atoms of another type. The atomic weights which had hitherto been found were actually mean values, arising from the fact that a certain quantity of one kind of atom had become mixed with atoms of another kind.

The Properties of Isotopes. We may now inquire how these two kinds of atoms will behave with respect to each other, and why they are now considered to be two different substances, especially when the atomic weights of the various elements have been the basis from which a classification of the elements has been made. These different kinds of atoms are found to be identical in their chemical properties; the one forms the same chemical compounds as the other, and it is precipitated from a solution by the same reagents as precipitate the other. In short, we may say that when chemistry is asked to decide the nature of these substances, it says that they are one and the same substance. Finally, the light which these substances emit when placed in an electric arc may be examined. Amongst all the known elements there are no two elements which give the same spectrum—the wave-lengths of the light and X-ray spectra are characteristic of the particular element used, so that they can be used as a reliable means for purposes of identification. The spectra of the different kinds of atoms of the same element, however, exhibit light and X-ray spectra which are absolutely identical. In general, we have to recognize a substance by its properties, so that under these circumstances we should have to say that these substances are one and the same. Hence the reason why these bodies had not been separated was due to the fact that they behave similarly with respect to all other bodies or influences. Whatever causes one part to dissolve, or be precipitated, will also cause the other part to behave in an identical manner. Hence, whenever the two components have been mixed, we have no way by which they can again be separated.

How can these substances be arranged according to the periodic system of the elements? The number of places available in the periodic table for the different elements is definitely fixed—from hydrogen to uranium there are ninety-two such places. For a whole series of new substances there are no vacant places. If, however, we arrange the various substances according to their chemical properties, then these substances cannot be put into new positions, for, having the same chemical properties, they must be placed together in one and the same place. It is quite certain that this system of classification would have to undergo a radical change if several substances were placed in one position. These different substances having the same chemical characteristics are called '**isotopes**'. They are only different from one another by the fact that their atomic weights are different—only a little, but yet there is a decided difference. This indicates to us the manner in which it may be possible to separate isotopes —at least to separate them so far that the two isotopes are obtained in different proportions in two different fractions of a given supply. Such a separation is only possible if use is made of the fact that there is a small difference in their atomic weights. This is exactly what Aston was able to do in the above apparatus, for the deflection of each individual particle depends only upon its atomic weight ; it is quite independent of all its other chemical properties. Attempts have also been made to separate the isotopes by diffusion, since the velocity of diffusion depends upon the masses of the particles. The experiments were extraordinarily difficult and success was only possible after a long series of carefully planned experiments had been made ; finally, however, results were obtained which completely confirm the experiments with the mass-spectrograph.

The Atomic Weights of Isotopes. We have not yet quite exhausted the results which can be obtained from the experiments made by Aston. The mass-spectrograph can be calibrated by using in it gases of known atomic weight, and which only produce one image on the photographic plate. For example, the positions of the images when helium (4), oxygen (16) and carbon di-oxide (44) are used can be employed for the purposes of calibration. Having done this, interpolation can be resorted to, in order to fix the atomic weights corresponding to the isotopes which give images in other positions. In a very carefully performed series of experiments Aston investigated all possible elements and measured the distances of the images from the central position P. For a whole series of substances which had hitherto been regarded as pure elements he found that they were really mixed elements consisting of two or more isotopes, and when the atomic weights of these isotopes were calculated, a very surprising result was obtained. All those elements having a fractional atomic weight were found to be mixtures of two or more isotopes. On the other hand, all elements having an integral atomic weight were found to be pure elements and not mixtures ($C = 12$, $O = 16$, $He = 4$). The fractional numbers which had hitherto represented atomic weights were really mean values. This was not the only remarkable result discovered, for it was found that the **atomic weights of the isotopes themselves were always integral.** The inert gas neon, for example, having a mean atomic weight 20·2, was shown to consist of two isotopes having atomic weights 20 and 22 respectively. For chlorine, 35·46, Aston found two isotopes, the atomic weights of which were 35 and 37. In the table of atomic weights given on p. 108, the atomic weights of all the known isotopes have been recorded. These numbers are the latest values published by Aston in 1927, and the order in which they are arranged indicates the prevalence with which they occur.

This discovery takes us back to the early days in the history of the atomic theory, when the fact that the atomic weights were represented by fractional numbers played an important rôle in developing that theory. From the fact that many of the elements had an integral atomic weight **Prout** concluded that all atoms were built up of hydrogen. This hypothesis had to be discarded, however, when it was found that there were many elements for which the atomic weight was represented by an integer plus a fraction, *e.g.* $Cl = 35{\cdot}46$. We now see that these fractional values are not really genuine, but they are the mean values of two or more integers which represent the atomic weights of their isotopes ; thus, the difficulty which was then urged against Prout's hypothesis disappears. In fact, the rudiments of this assumption have received additional support. The number of elements known is much greater now than in those days, in fact it is far greater than they ever dared to imagine, and yet amongst all their isotopes there is not one which has an atomic weight which is not a simple multiple of that of hydrogen. For the present it must suffice if our ideas concerning atomic weights are accordingly changed a little. For our further investigations, however, we have a clear hint that we may expect to obtain many fundamental ideas concerning the structure of the atom. We shall learn more about this from the pages which follow.

PART III

THE STRUCTURE OF THE ATOM

INTRODUCTION

THE arguments which have been previously advanced have shown that the whole material world is built up of molecules and atoms ; we have also learnt from them the more important laws under which these small particles move and have discovered something about their size and number. The ascertaining of this number and with it the mass of the individual atoms must be regarded as the crowning point in this field of inquiry.

Although the atom is so extraordinarily small that one can never hope to see it even with the aid of the most powerful of microscopes, yet man's inquisitive nature has led him to make further investigations. What are the atoms, can they be recognized by means of their shape, are these small particles the smallest entities which exist, or can the atoms be further divided ? And if we do succeed in discovering that atoms are composite bodies, can we say how these components are arranged in the atom ? Can the observed properties of chemical compounds, the emission of light, and their electrical properties be attributed to any one portion of the atom ? The mere fact that we are able to propound such definite questions indicates, at least, that we have been able to answer many of them.

Several incentives to the execution of such researches might have been found from the results which had already been obtained. The fact that the elements had been classified according to the Periodic System suggested that some sort of relationship might exist between the various atoms. It would be very improbable that the elements should be able to be classified according to their chemical properties, if such a relationship did not subsist. At that time, however, there were no means whereby the problem could be tackled, but later on, several vantage points were gained, from which one could set out to investigate the mysteries of the realm of the infinitely small. As we shall see presently, these points were not all found in the science of chemistry : it was only when research workers had ceased, at least in part, to pay strict attention to the sections into which science had been divided, that progress became possible. Nature does not recognize chemistry and physics, astronomy and geology, but only bodies which always assert themselves together with all their properties. Much advantage has always been gained when the results obtained in one field of research have become known and utilized in another, and very often much important progress has been made in one branch of science when someone has been discovered who had a thorough knowledge of several branches of science. In connexion with the present problem, we may safely say that such progress would never have been possible if only the methods of chemistry had been used, for it was by means of physical methods, applied to chemical problems, that a solution was eventually obtained.

CHAPTER I

MAGNETISM

1. THE FUNDAMENTAL LAWS OF MAGNETISM

FUNDAMENTAL PHENOMENA. It is quite certain that the compass had been used in Europe about the year 1200. At that time it was also known that a magnetized needle pointed to the north ; it also appears to be very probable that the Chinese were acquainted with the compass and had used it much earlier, probably a thousand years. The fact that a naturally occurring magnet, known as lodestone, could attract pieces of iron was known about the beginning of the Christian era. A few persons had reported the fact that repulsion had been occasionally observed. It is not easy, however, to form a reliable estimate of the view held by these ancient philosophers, since **Pliny**, for example, asserts that a diamond completely robbed a magnet of its magnetic properties.

We are indebted to an English physician, **William Gilbert** (1540–1603), for our first reliable ideas concerning magnetism. In a treatise on magnetism (London, 1600) he explained the fact that a magnet always pointed to the north on the grounds that the earth itself was a very large magnet. He knew that there were two kinds of magnetism, and that they produced effects opposite in character. Gilbert appreciated the facts that like poles repel one another, whereas unlike poles attract one another, and that a piece of soft iron when held near to a magnet itself became a magnet. He knew also that the magnetism of a steel bar could be destroyed by heating it to the temperature of red heat, and that when a magnet had been broken into smaller fragments, each fragment remained a small but perfect magnet.

The most important of all these discoveries is the fact that when a piece of iron is brought into a magnetic field it becomes a magnet. This can very easily be shown, for so long as a piece of soft iron remains in the neighbourhood of a magnet it has the power of attracting other small pieces of iron to itself. As soon as the attracting magnet is removed, the pieces of iron which it has attracted fall away from it. The axioms concerning the mutual effect of two bodies on one another which were mentioned on p. 36 *et. seq.* must now be recalled. We there learnt that two bodies can only produce an effect on each other, when the force which originates in each is identical in character. Here a violation of this axiom would have been found if the magnet had been able to attract an unmagnetized piece of iron. We see, however, that this is not so. The magnet first causes the soft iron to become magnetized, mutual attraction between the two magnets then taking place. The fact that Gilbert had explained the tendency which all magnets had to seek the north, by postulating the existence of terrestrial magnetism, may mean that he was acquainted, perhaps unconsciously, with this axiom. The magnetism which is produced when iron merely approaches another magnet is said to be **induced magnetism.**

The other observations, viz. that when a magnet is broken into fragments each small piece remains a magnet, became very important at a later date, when an attempt was made to explain the phenomena of magnetism.

Coulomb's Law. In the two centuries which followed, little progress was made. The question was first tackled really scientifically by the French physicist **Coulomb** (1736–1806), who, in 1785, began to measure the forces of attraction and repulsion between magnets. If a magnet is obtained in the form of a long thin needle it is very easy to show that the magnetic forces are especially strong in the neighbourhood of two points situated near the ends of the needle; they are called the **poles of the magnet.** The two poles of a magnet always exhibit the two opposite kinds of magnetism—one at each end. In general, an experiment arranged to detect the effect of one pole is disturbed by the presence of the second pole. Coulomb procured some very long needles so that the disturbing pole was so far removed that its effect could be neglected. The one magnet was suspended by means of a brass wire from the head of a torsion balance, the force between the two poles N and N_1 being determined by the torsion of the wire necessary to hold the magnet in its zero position. He found, first of all, that the force was inversely proportional to the square of the distance. Qualitatively he was able to show quite easily that some poles produced an effect equal to that of several others placed together; *i.e.* that the effect was intensified not only when one of two attracting poles was increased, but also when the other was increased. He established the law, therefore, that the effect of two poles on each other is directly proportional to the product of their pole strengths and inversely proportional to the square of their distance apart, *i.e.*,

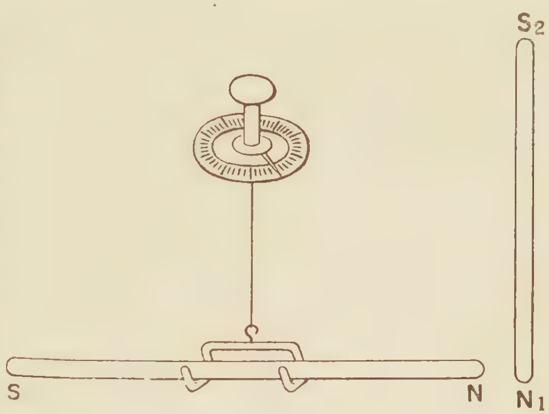


Fig. 94.

$$f = C\frac{mm'}{r^2} \qquad . \quad . \quad . \quad . \quad . \quad (95)$$

The justification for saying that the cause of the difference in the magnitude of the attraction (or repulsion) is due to the poles is furnished by the following facts. First, if two magnets act upon a third (this last one may be the magnet used in a torsion balance) and produce effects which, when the magnets are kept at the same distance apart, are in the ratio of 1 : 3, then the effects produced when the magnets are placed at any other distance apart will always bear the ratio 1 : 3 to one another. Secondly, if the same two magnets are compared by means of another torsion balance the effects which they exert upon its magnet will again be in this same ratio of 1 : 3. It is therefore necessary to conclude that this ratio 1 : 3 is due to the strengths of the two magnets. Coulomb said that each magnet had a definite **pole strength** which he denoted by m, and then expressed his results in the form of the above law. It is at once seen that equation

(95) is exactly similar in form to that which expresses Newton's law of gravitational attraction. In the present instance, however, the force may be either one of attraction or of repulsion, whereas gravitational forces are always forces of attraction. Later on the law was subjected to a more critical examination, but it was always verified by the results which were obtained.

Magnetic Units. In 1832 the mathematician, **Karl Friedrich Gauss,** used the above law in order to determine the unit of magnetic pole strength, m. He saw that the constant C of equation (95) would become equal to one, if a **unit pole** were defined, **so that when it acted upon another pole of similar strength at distance 1 cm. away, the force of repulsion was 1 dyne.** By this means, Coulomb's law can be written more simply in the form

$$f = \frac{mm'}{r^2} \quad . \quad . \quad . \quad . \quad . \quad (96)$$

Historically, the determination of this unit was of very great importance. It was the first time that a quantity (*i.e.* the magnetic pole strength) which was not derived in the ordinary course of mechanics, was measured in terms of the units used in that branch of science, viz. the dyne and centimetre. The absolute system of units was here really employed for the first time. The extension of the same methods to electrical quantities no longer presented any difficulties. Gauss, himself, points out this fact in his work on the theory of potential published in the year 1843. Since that time the determination of electrical and magnetic units in terms of mechanical units has been made use of in scientific work; electricity, too, plays such an important rôle in daily life that it has been necessary to specify its units in terms of those of mass, length and time.

As we saw earlier on, the dependence of one unit upon others can be represented by means of a dimensional equation, which expresses this relationship explicitly. This dimensional equation is obtained from the one which is used to define the unit; in the present problem it is equation (96). Since the dimensions of a force are $f = [mlt^{-2}]$, we have

$$[mlt^{-2}] = [\boldsymbol{m}^2 l^{-2}],^*$$

so that the dimensions of pole strength are given by

$$[\boldsymbol{m}] = [m^{\frac{1}{2}} l^{\frac{3}{2}} t^{-1}] \quad . \quad . \quad . \quad . \quad (97)$$

Magnetic Potential. From Coulomb's law of force the theory of magnetic potential follows logically—just as the theory of gravitational potential was developed from Newton's law of gravitation. The term m, however, no longer represents an inert mass, but a quantity of magnetism or a 'magnetic mass' called the pole strength [*cf.* p. 56 *et seq.*].

The magnetic potential at a distance r from a pole is expressed by $U = m/r$, and the work which must be done in order to bring unit quantity of magnetism from a point at a distance r_1 to a point r_2 is $U_1 - U_2$, where U_1 and U_2 are the magnetic potentials at r_1 and r_2 respectively. The amount of this work is independent of the actual path traversed or the actual position finally reached on the equipotential surface.

The law of magnetic force can therefore be expressed somewhat differently as follows. The force on a pole of unit strength at a distance r is equal to

---

* When it is necessary to make a distinction between m, the mass of body, and $\boldsymbol{m}$, the pole strength of a magnet, the two types here indicated will be used.

$\frac{m}{r^2}$. This force is called the **intensity of the field** at distance r and is denoted by H, so that the force on a pole of strength m' is $m'H$. If now the above value of H is inserted in this equation we have Coulomb's law expressed in its original form. This new way of expressing the law possesses the advantage that it remains valid even when several poles, and not one, are used to produce the field.

Magnetic Moment, Intensity of Magnetization. At a large distance from a pole the intensity of the magnetic field can be considered constant over a small region. The earth's magnetic field will therefore exert equal forces upon the poles of a magnet, which is only a few centimetres in length; the one pole will be repelled by a force which is equal to the force by which the other is attracted. Thus when a small magnet is also free to move, it will tend to move as a whole neither towards the north nor towards the south pole of the earth. But if the line joining its pole, *i.e.* its magnetic axis, Fig. 95, makes an angle with the lines of magnetic force, *e.g.* it may be perpendicular to them, these forces will cause it to rotate until the magnetic axis lies along the lines of force in the field. The magnitude of the two forces is $+mH$, and since they each act at a distance l from the point of support, the total moment of the forces is $2mlH$. [The distance between the poles of the small magnet is $2l$.] m and l are characteristics of the magnet, viz. the strength of its poles and their semi-distance apart. The quantity $2ml$ occurs so frequently in the study of magnetism that it has been denoted by M, and is called the **magnetic moment** of the magnet. The magnetic moment per unit volume is called the **intensity of magnetization,** I, so that

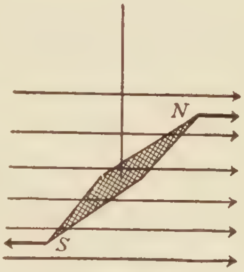


FIG. 95.

$$I = \frac{M}{V} = \frac{m}{A},$$

where V is the volume, and A the cross-sectional area of the magnet.

Along any short distance the intensity of the magnetic field can always be considered constant. If such a short element of distance is denoted by dl, then the work done in moving a unit pole against the magnetic force is $H.dl$ and this is equal to the small difference of potential dU between the extremities of the element dl; hence $H.dl = dU$, or $H = \frac{dU}{dl}$. If the mean value of the magnetic intensity over a distance l is H, then the work done is $H.l$. This work measures the difference in magnetic potential between the two extremities of the distance l.

Influence of the Medium separating two Magnets. When we consider Coulomb's law, as represented by equation (96), it may occur to us that it merely contains the distance apart of the two poles, without indicating or assuming anything with respect to the nature of the intervening medium. It has therefore been said frequently, in fact even until quite recently, that Coulomb's law, and Newton's law, to which it is cognate, is purely a law of action at a distance, *i.e.* the forces of gravitation and magnetism arise without the intervention of a medium. This, however, is not a logical deduction from the law. The laws of Coulomb and Newton are merely an expression of facts; in particular, the law of Coulomb only asserts what he had discovered experimentally with his torsion balance. As an expression of such facts, these laws are free from any hypothesis whatsoever. On the contrary, however, every hypothesis, whether it postulates the idea of action of a distance or not, must agree with this law, for otherwise it would contradict the facts and therefore be false. Moreover, as already mentioned on page 60, Newton had expressed the opinion that direct action at a distance was impossible. It is quite certain that, at that time, nothing was known concerning the action of the field in connexion with gravitational or magnetic forces at a distance, and, moreover, at that time several physicists assumed that action at a distance was possible. Newton, however, does not belong to them.

M. Faraday (1791–1867) certainly did not agree with such views. He made the daring resolution to prove experimentally that magnetic actions were made possible by the aid of the intervening medium (electrical forces were similarly investigated).

With rare sagacity he devised suitable experiments to detect the influence of the field, and by performing these experiments with such wonderful skill, he rendered the greatest possible service to the science of magnetism and electricity, which was then quite young.

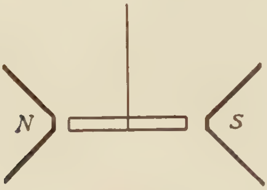


Fig. 96.

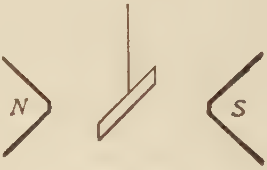


Fig. 97.

In one of his books Faraday wrote: 'For my part, when I consider the behaviour of a vacuum towards magnetic forces, and the general nature of magnetic phenomena outside the magnet, I am inclined rather to the idea that the propagation of the force depends upon some such process (in the medium) rather than that attraction and repulsion are the results of action at a distance.' In the long run this method of representation most nearly approaches the true condition of things in Nature. Whilst Faraday was investigating the properties of as many substances as possible in the magnetic field, he discovered that there were two types of substance. If a long rod of the material under examination is suspended by means of a fine thread between the tapering poles of a strong magnet, it is found that iron, nickel, and several other substances, arrange themselves so that their magnetic axes lie along the line joining the poles; *i.e.* they set parallel to the lines of force, Fig. 96; on the other hand, bismuth and copper arrange themselves perpendicular to the direction of the lines of force, Fig. 97. If they are displaced from this position they return to it on their own account, and in so doing execute oscillations about this direction as the zero position, or position of equilibrium. Faraday described bodies of the first type as being **paramagnetic**; those of the second type were **diamagnetic**. Paramagnetic bodies are attracted by a single magnetic pole, but diamagnetic bodies are repelled. He also showed that the air, oxygen, and several other gases, were paramagnetic. The chief discovery which he made in this field of knowledge was that it depends upon the medium between two magnetic poles as to whether a body will behave paramagnetically or diamagnetically. If the body is more strongly paramagnetic than the medium, then it arranges itself along the lines of force just like a paramagnetic body; if, however, it is less paramagnetic than the medium, it places itself across the lines of force and so appears to be diamagnetic. Hence it follows that in one medium a body may appear as a paramagnetic one, whilst, in another, it may appear as a diamagnetic body.

The repulsion of paramagnetic bodies in a more paramagnetic medium is therefore not an original phenomenon; it depends rather on the circumstance that the surrounding medium is more strongly attracted than the paramagnetic body. It therefore tends to move away from the magnet just as a piece of wood in water ascends to the surface after having been depressed; not because the earth repels it, but because the water is drawn more strongly downwards.

In this way it was shown that the medium does play a part in the phenomena of magnetism.

Faraday's Lines of Force. These reflexions of Faraday on the effect of the medium and the propagation of magnetic force from point to point in a medium, enabled him to express the same ideas by assuming the existence of lines of force. He represented the forces between two poles by means of lines drawn from pole to pole; these lines indicated everywhere the direction of the magnetic intensity at the point in question. In these lines, however, he saw much more than the mere representation of the direction of the magnetic force at any point; to him these lines were like pieces of indiarubber stretched between the two poles. The tendency to contract, which such bands would have, represented the action of the magnetic forces. The greater the force of attraction the more closely did these threads lie together. These directional lines are very easily rendered visible. A magnet is placed under a glass plate which is then covered with iron filings. The iron filings arrange themselves along the lines of force when the plate is gently tapped; a picture like that shown in Fig. 98 is obtained.

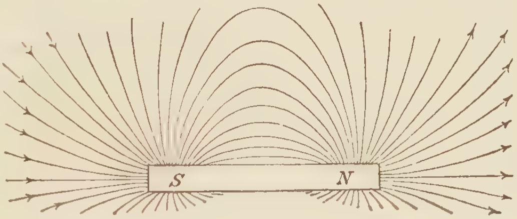


FIG. 98.

When magnetic forces are represented by means of these lines many persons are enabled to form a better conception of the field; they have therefore been so far developed that, at the same time, they represent the magnitude of the magnetic intensity at every point. Let us first of all fix our attention on a single pole, *e.g.* the pole of a long magnetized needle which is perpendicular to the plane of the paper. Then all the lines of force in this plane will be straight lines proceeding from the pole P. In order to represent the field strength at any distance r from the pole by means of lines, we describe, with P as centre, a sphere of radius r. The field strength at all points on this sphere will be the same. Suppose now that the surface of this sphere is divided into small portions, each 1 cm.^{2} in area, and that through each 1 cm.^{2} so many lines of force are drawn as will represent the field strength at that point. Then the field strength is equal to the number of lines of force per cm.^{2}; the intensity of the field will therefore be greater when the density of the lines at the point in question becomes greater. In the present example, if the pole strength is P, the number of lines per cm.^{2} is P/r^2. Since the total surface area of the whole sphere is $4\pi r^2$, it follows that the total number of lines emerging from the pole P is

$$4\pi r^2 . P/r^2 = 4\pi P.$$

It is at once seen that this number is independent of the radius r. The lines therefore emerge from P and every surface completely enclosing P is cut by $4\pi P$ lines of force, as is shown in Fig. 99.

The unit of field strength has been called a 'gauss' in honour of the mathematician **Gauss**. With the aid of strong artificial magnets fields up to 50,000 gauss have been obtained. This means that a unit magnetic pole placed in this field will be attracted or repelled by a force of 50,000 dynes.

Magnetic Induction. We have seen that all bodies which are acted upon by a magnet become magnets themselves. Speaking in a strictly logical way, we have to say that the air itself shows magnetic properties when it is traversed by magnetic lines of force. In this fact we have direct evidence of the action of the medium in propagating magnetic disturbances. Since these disturbances can be transmitted through empty space, the æther itself must be regarded as capable of becoming magnetized. This property of the æther is the basis of **Maxwell's** electromagnetic theory of light. In the space between the two poles of a magnet we have a field intensity H. It is equal to the number of lines of forces passing through a unit area of cross section. If a body is brought into this field it will in general become a magnet. Its intensity of magnetization is I. The intensity of magnetization produced by a unit field, *i.e.* $H = 1$, is called the magnetic susceptibility κ. We therefore have $\kappa = I/H$, or $I = \kappa H$.

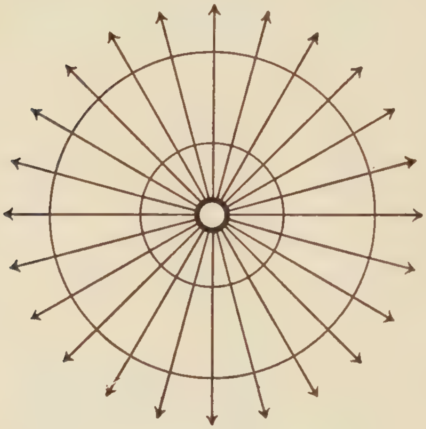


FIG. 99.

The quantity κ is a characteristic of the substance under examination. For paramagnetic bodies it is smaller than 10^{-4} but positive, *i.e.* the excited magnetism lies in the same direction as that of the magnetizing field. For diamagnetic bodies κ is negative, its order of magnitude being 10^{-6}. In the case of ferro-magnetic bodies κ can assume large values; its value may even rise to beyond 1000. The value of κ depends upon the value of the field strength, as we shall shortly prove. For the methods of determining H and I reference must be made to more detailed treatises dealing with the subject.

If the field strength is gradually increased, the magnetization does not pass beyond a certain value I_0. When in this condition the body is said to be 'saturated'.

The foregoing remarks really embody the laws of magnetic phenomena. These laws and the inferences which are deduced from them permit us to describe, in many respects, all the known facts. In particular, where it is only a question of the magnitude of the forces, and no attempt is made to explain them in terms of any theory concerning the nature of magnetism, this method of presentation suffices for the needs of industry. This is called the phenomenological method of representation, *i.e.* the laws are merely descriptive of the phenomena in question.

2. THEORIES OF MAGNETISM

Man, however, is not satisfied entirely with this method of presenting things. Even if, at times, some satisfaction was gained when it was found that some calculations agreed with the reality, yet the desire for an explanation of the why and wherefore was ever paramount. There have always been some of the best thinkers who were not satisfied with a mere numerical agreement. They wished to learn how any particular property depended upon those of the ultimate constituents of matter. An atomistic explanation of the phenomena was therefore urged in place of the phenomenological one. The justification for this demand is perhaps best shown by the fact that it was the more gratefully accepted and the more readily suggested, the better one was in a position, in the course of scientific research, to give explanations more in complete accord with the observed relations.

The Theory of Magnetic Fluids. The oldest theory of magnetism was that which assumed the existence of two magnetic fluids ; this theory was obviously formed, not in conjunction with the known facts, but rather after the idea of phlogiston and the electric fluids. The two fluids were said to be spread throughout the iron. When they were thoroughly mixed, the rod of iron was unmagnetized ; if one fluid collected at the one pole, and the other at the second pole, then the iron became a magnet. This hypothesis, however, is not consistent with Gilbert's discovery that when a magnet is broken in two one does not obtain two separate magnets each with north- or south-seeking magnetism only, but two perfect magnets each with a north and south pole ; the two new poles appear at the place where the original magnet was fractured.

In order to retain the hypothesis it was assumed that the two fluids were not free to move throughout the magnet but attached to the particles which constitute the magnet ; the fluids were able to move, however, a little distance around each particle. In this way the two fluids will be separated from one another in each particle, *i.e.* the individual molecules become little magnets. When all these are put together opposite kinds of magnetism are always to be found as neighbours, so that all effects are neutralized.

In this form the explanation contains two assumptions : first it assumes the presence of small elementary magnets which are said to explain the phenomena that magnetic forces can only be noticed at the poles ; in the second place the presence of these molecular magnets is explained by assuming the presence of a magnetic fluid which is mobile. Now the first assumption is strongly supported by observations, but the second is quite arbitrary and contradictory to experience, so that it becomes untenable. The abandonment of this theory was hastened by the fact that the idea of ‘ caloric ’ had been discarded in the science of heat. The important fact, which the hypothesis of magnetic fluids could not accommodate, was that a piece of iron cannot be magnetized to an unlimited extent, but only to a certain degree. When this limit has been reached all further endeavours to increase the amount of magnetization in it fail. If, however, both the liquids could be separated, one cannot see why this separation should not be increased when the magnetizing force becomes greater, but should have a definite limit. The hypothesis which assumed the existence of magnetic fluids has been discarded, but the first idea, viz the existence of small magnets, has led to the theory which we must now discuss.

The Theory of Molecular Magnets. It has therefore been assumed that the magnetization of iron and kindred bodies is an atomic property, *i.e.* the atoms themselves are all small magnets, each with a north- and south-seeking pole. Under ordinary circumstances these small magnets are arranged irregularly throughout the material, so that the effect, for example, of a north-seeking pole is always counterbalanced by that of a south-seeking pole situated near to it. The process of magnetization therefore consists in rotating these molecular magnets so that they point in a definite direction. Under the influence of small magnetizing forces this rotation is not complete; when, however, the small magnets have been rotated so that they lie in the direction of the magnetizing force, the amount of rotation cannot be increased. The phenomenon of magnetic saturation is therefore easy to explain. This theory has been developed mainly by **Weber**. He computed the relationship between the strength of the magnetizing field and the amount of magnetism produced, but he was not able to calculate the new effects which were continually being discovered. It soon appeared that **ferromagnetism**, on the one hand, and **para-** and **diamagnetism**, on the other, were really two very different phenomena, so that we shall now discuss them separately. Although there are paramagnetic bodies which can be magnetized very considerably, there are no diamagnetic bodies which can be magnetized to such an extent—diamagnetic substances are always weak. We therefore differentiate between three types of magnetic bodies. First, we have the strongly paramagnetic bodies which are the elements comprising the iron group of elements; secondly, we have the feebly paramagnetic substances; and thirdly, the (feebly) diamagnetic bodies.

Iron, nickel and cobalt belong to this first group of elements. Aqueous solutions of iron compounds are also strongly paramagnetic.

Aluminium, magnesium, palladium, platinum, oxygen and nitrogen are bodies which are feebly paramagnetic.

All other gases, together with bismuth, water, zinc, silver, mercury, copper, lead, and other substances, are diamagnetic.

Some very important facts for the theory of magnetism were revealed by the discovery made by **F. Heusler** at a later date (1898). He found that alloys formed from certain metals, which were only very feebly magnetic themselves, could be magnetized to almost the same degree as iron itself. The alloys of manganese, and especially the alloys of manganese-aluminium-copper, exhibit these properties.

Ferromagnetism. When a piece of unmagnetized iron is brought into a magnetic field and the degree of magnetization measured with respect to the strength of that field, the following facts are observed. Initially, when the magnetic field is very weak, the magnetization is proportional to the field strength; this is followed by a region in which the magnetization grows rapidly, after which it can only be increased with difficulty and finally not at all, however much the field is increased (magnetic saturation has been reached). If, in Fig. 100, the abscissae represent the strength of the magnetic field, whilst the ordinates give the intensity of magnetization, the curve *OA* represents the relationship between these two quantities.

A very small portion of the curve near to O is very nearly a straight line; this is due to the fact that the magnetization is proportional to the strength of the magnetizing field. Then the curve turns sharply upwards and the intensity of magnetization is very easily increased. Soon afterwards magnetic saturation is reached and the curve runs horizontally. Weber's theory of magnetism was unable to account for the facts indicated by the middle portion of the curve.

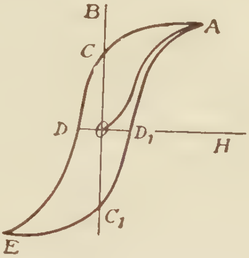


FIG. 100.

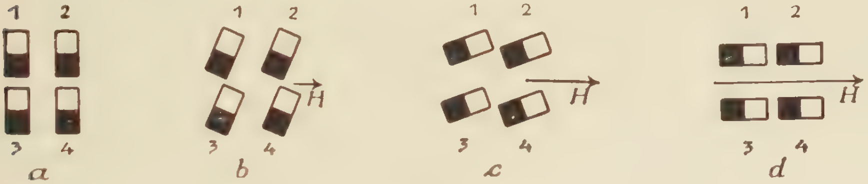


FIG. 101.

This difficulty has been removed by **Ewing**. He assumes that iron and similar elements are made up of elementary magnets, each consisting of a group of atoms so that each elementary magnet consists of a great number of atoms (since then it has been found that quite small crystals of iron form these elementary magnets). The elementary magnets themselves consist of molecular magnets all having the same orientation. The elementary magnets which lie near to one another in a piece of unmagnetized iron have different orientations in general; they are so arranged that one pole is near to another of the opposite kind but belonging, of course, to another elementary magnet. The arrangement accounts for the unmagnetized condition of a large piece of iron. If an externally applied magnetic field tends to rotate the elementary magnets, then, so long as the field is weak, neighbouring poles must exert a force upon the molecular magnets which cannot be neglected in comparison with the external forces. Let Fig. 101 represent an elementary magnet which is composed of four molecular magnets. Let H, the external field, be perpendicular to the direction of the original magnetization [Fig. 101 (*a*)]. The small magnets will endeavour to align themselves in this direction [*cf.* Fig. 95]. In so doing, however, the small magnets will experience forces tending to cause them to return to their original direction. These forces will be due to the opposite kinds of magnetism in the neighbouring poles of such pairs of magnets as 1 and 3, 2 and 4. The amount of rotation will be greater when H increases, but as soon as the magnetizing field is removed the small magnets will return to their original positions. If the field is increased, the distance between the opposite poles which were originally near together, will become greater, the small magnets tending to take up a horizontal position. The magnetic forces between 1 and 3, 2 and 4, will be less, but those between 1 and 2, 3 and 4, will be greater. Finally, these two forces will become equal so that the small magnets are in a state of unstable equilibrium when the least increase in the field strength suffices to move them into the positions shown in Fig. 101 (*c*). The passage from (*b*) to (*c*) corresponds to the central portion of the curve OA in Fig. 100. The particles are now able to rotate completely into the direction of the magnetizing field, *i.e.* the state of magnetic saturation is reached [Fig. 101 (*d*)].

Ewing's assumption has proved itself to be sufficient to explain the subsequent behaviour of magnetic bodies. When the field strength H is reduced from its maximum value only a portion of the magnetism in the iron is lost, so that even when the field is zero there is, in general, a considerable amount of magnetism retained by the iron (in Fig. 100 this is represented by the intercept OC). This is called the **remanent magnetism,** or **remanence.** In this state the small magnets which are present in the iron remain fixed in their new position, the opposite poles of the different molecular magnets being adjacent. To destroy the magnetism completely a fairly large magnetic field, acting in the negative direction, must be applied. In Fig. 100 this occurs when $H = -OD$. This counter force measures the force with which the magnet holds its magnetism. It is therefore called the **coercive force.** If the field is increased further in the negative direction the iron becomes magnetized with reversed polarity. The magnetization again increases rapidly and reaches at E the same saturation value as it did at A. When the field strength is again reduced the change in magnetization is represented by the curve EC_1D_1A. This curve is like the portion $ACDE$ in all respects except in that of sign.

By striking the iron, or disturbing it mechanically in some other way, the remanence of a magnet is decreased ; it is also diminished by heating. At certain definite temperatures the remanent magnetism vanishes completely ; for iron, cobalt, and nickel, these temperatures are respectively 769° *C.*, 1075° *C.* and 356°. At still higher temperatures iron and steel behave exactly like a paramagnetic body.

Para- and Diamagnetism. From the above remarks it will be seen that ferromagnetic phenomena can, in general, be attributed to the properties of the molecules or atoms. The explanation of para- and diamagnetic phenomena is also based on the properties of atoms and molecules. Twice in the history of science have our ideas concerning magnetism been not upset but, in general, extended and deepened, by discoveries made in other branches of physics ; these discoveries have thrown new light on the subject. The further treatment of magnetism will therefore be postponed until we have learnt the value of these discoveries.

CHAPTER II

ELECTROSTATICS

1. Fundamental Phenomena

1. THE only fact about electricity known to the ancients was that a piece of amber, rubbed with wool, acquired the property of attracting light bodies (and repelling them after contact had occurred).

2. There are conductors and non-conductors of electricity ; or rather, there are good and bad conductors of electricity. In practice all intermediate grades are found. The best conductors are the metals, especially silver and copper ; then come the aqueous solutions of acids, bases and salts ; bad conductors or insulators are such substances as wood, paper, straw, glass, paraffin wax, ebonite, fused quartz and amber. Water is a semi-conductor, so that the last-named substances are only insulators when they are free from moisture.

3. There are two, and only two, kinds of electricity—positive and negative. They have been so named since when equal quantities are placed upon the same body that body remains unelectrified. Whether we have, in these two kinds of electricity, two different entities producing effects which are opposite in character, or whether we have only one, is a problem concerning which there has been much lively discussion. If there is only one sort, then an excess of it will be positive, whilst a deficiency is negative (this is somewhat similar to two vessels one of which contains compressed air and the other rarefied air). Even to this day the above question cannot be regarded as having been solved. Now according to the rules of a rational science two things ought not to be assumed when one is sufficient to explain the phenomena in question. For this reason, during the last century, one has favoured the theory which postulates the existence of one electric fluid, but the results of modern work have many times suggested that this theory is not true. The question must therefore be considered as still unsolved.

4. Similarly charged bodies repel one another, whereas bodies with unlike charges attract one another. All these laws are easily corroborated by means of experiments.

When these properties of electric charges had become known, instruments were soon devised in order to detect the presence of electricity. In them, light bodies, suspended from one end, are used ; the motion of the other end is then easily observed and measured. Formerly one or two very thin leaves of gold or aluminium were used. These leaves were either repelled by a metallic conductor or else attracted to an oppositely charged body. Nowadays very small metallic strips are used ; in the precision instruments fine platinum wires are often employed. Fig. 102 shows the essential parts of some of the more usual types of these instruments ; (*a*) is the aluminium-leaf electroscope due to **Exner**, (*b*) is **Braun's** pointer-electrometer, (*c*) is **Wulf's** universal electroscope, and (*d*) is **Wulf's** double-string electrometer.

The Conservation of Electricity. When two bodies are rubbed together both of them always become electrified, *i.e.* they acquire the property of attracting other light bodies. Exact experiments have shown that the one is always electrified positively whilst the other becomes negatively electrified ; the two sorts are equal in amount and when allowed to come together again they completely neutralize one another. It is therefore necessary to conceive that the process of electrifying bodies by friction is not one in which electricity is created but only that the two sorts of electricity, which are always present in a body, have become separated, or that *one* electric-fluid has passed from one body to the other. It may here be remarked that this assumption has always been confirmed by all later experiments, whatever their nature may have been. In no experiment is positive or negative electricity alone produced. This fact may be better expressed as follows : whenever electricity of one type appears at some place **an equal amount of the other type of electricity can always be found at some other place.** We can therefore place the law of the conservation of electricity at the side of those of the conservation of mass and of energy, and, as with these, no exception to it has ever been discovered. This fact leads us to suppose that electricity, like matter, is a substance which remains unchanged irrespective of the phenomena in which it may appear.

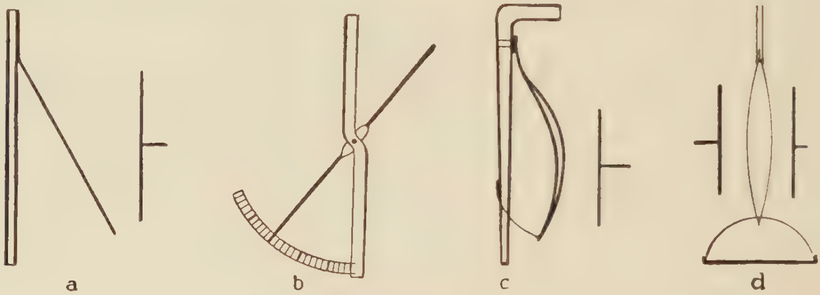


FIG. 102.

The question may possibly now be asked, why have we not allowed energy to be regarded as a ‘ substance ’ when it behaves in a similar manner ? (p. 184). The answer is as follows : in the case of energy the phenomena were not at all the same. Since energy can appear in various forms, kinetic, elastic, electric, electrokinetic, and in several other forms as radiation, it would be difficult to imagine that one and the same entity was responsible for them all. In addition, the principle of the conservation of energy was in some instances (when it appeared as mechanical, kinetic, or potential energy) a necessary consequence of the laws of force, especially those concerning action and reaction [*cf.* p. 34 *et seq.*], and independent of any assumption concerning the nature of the energy. Finally it must be noted that this does not say that energy is not a substance, but only that the above phenomena do not prove, or even indicate so emphatically, the conclusions we have adopted regarding mass and electricity.

Coulomb's Law. The first quantitative law concerning electricity which was derived from actual measurements was proposed by **Coulomb** in 1785. The law states that the force with which one electric charge q

attracts another charge q' (or repels a like charge) at a distance r away is

$$f = C\frac{qq'}{r^2} \qquad . \quad . \quad . \quad . \quad (101)$$

where C is a constant.

Coulomb discovered this law after performing experiments on electric charges with his torsion balance. His charges were measured in arbitrary units, but his law was used at a later date to define the unit of electric quantity; this was done by choosing for the unit of electricity that quantity which would make the above constant C equal to unity. **The unit of electric charge is therefore defined as that quantity which exerts upon a second equal quantity to it a repulsive force of one dyne when the distance apart of the two charges is 1 cm.** Coulomb's law can then be expressed more simply as

$$f = \frac{qq'}{r^2} \qquad . \quad . \quad . \quad . \quad . \quad (102)$$

Since the force f has dimensions $[mlt^{-2}]$, the dimensions of an electric charge, as found by applying the rules which have already been formulated, are

$$q = [m^{\frac{1}{2}}l^{\frac{3}{2}}t^{-1}] \qquad . \quad . \quad . \quad . \quad . \quad (103)$$

2. ELECTRIC POTENTIAL

The similarity between the above law and the corresponding one for magnetic forces and **Newton's** law of gravitational forces is at once very apparent. The potential theory which has been developed in connexion with these last two forces is therefore applicable to the electric field, which is defined as the space around a charged system in which electric forces are to be found. In order to bring a charged body nearer to a similarly charged body work must be done; if the charge on the stationary body is q, then the work, W, done in bringing up a unit charge from distance r_2 to distance r_1 is, as we have explained previously,

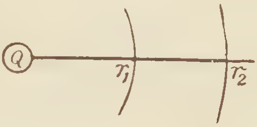


FIG. 103.

$$W = q\left(\frac{1}{r_1} - \frac{1}{r_2}\right) \qquad . \quad . \quad . \quad . \quad (104)$$

In this equation q appears instead of m since we are dealing with an electric charge instead of an inert mass.

When a unit of electric charge is brought up from infinity (for practical purposes we may say from such a distance that the electric forces are there almost zero) to the position at a distance r away, then the work done is q/r, which is defined as the **potential** due to the charge q at a distance r. Denoting this potential by V, we may therefore write

$$V = \frac{q}{r} \qquad . \quad . \quad . \quad . \quad . \quad (105)$$

If the unit of electricity is brought up to a sphere of radius r_0 carrying a charge q, then the work done is

$$V = q/r_0 \qquad . \quad . \quad . \quad . \quad . \quad (106)$$

which is the potential of the sphere. In bringing up a charge e the work done is

$$W = qe/r \qquad . \quad . \quad . \quad . \quad . \quad (107)$$

Capacity. The above formula (106) tells us that when the body carrying the charge has a definite size, then the charge upon it is proportional to its potential. If the potential is to be raised by an amount unity, then the same quantity of electricity, x, is always necessary, for from the equations

$$V + 1 = \frac{q + x}{r_0} \text{ and } V = \frac{q}{r_0}$$

we have, by subtraction,

$$x = r_0.$$

This means that in order to increase the potential by unity, a charge of r_0 units is required. In other words, in order to raise the potential of a large sphere, more electricity is necessary than for a smaller sphere, just as more water is necessary to raise by 1 cm. the level of water in a large vessel than when the vessel is small. One may therefore speak of the capacity which a body has for electricity. We have found that the capacity of a sphere is equal to its radius.

Now the idea of capacity can obviously be applied when the shape of the body is different from that of a sphere, for a definite quantity of electricity will be necessary in order to raise its potential by unity. If this quantity is C, then C is termed the **capacity** of the body. It is always possible to find a sphere which requires this same charge of electricity in order to raise its potential by unity. If the sphere has a radius r, then $C = r$. In order to raise by V units the potential of a body whose capacity is C, the amount of electricity, q, required is

$$q = VC \quad . \quad . \quad . \quad . \quad . \quad (108)$$

The Distribution of Electricity on Charged Bodies. The fact that two points are at the same potential implies that no work is necessary in order to move an electric charge from one to the other. This can only be so when the force on the charge between the points is zero. If therefore two points at the same potential are joined together by means of a conductor no charge moves along the conductor in either direction. If, however, the two points are at different potentials, then positive electricity will pass from the one at higher potential to the one at lower ; in so doing the place at the higher potential loses its charge so that its potential falls, whereas the place to which the charge flows gains both charge and potential. Hence, whenever an electric charge moves from one point to another, the potential at those two points will tend to become equal, and when these potentials are equal the motion will cease. When therefore a conductor is connected to another conductor the potential always assumes the same value at every point in them ; and if electricity is passing from a charged body to one which is uncharged the flow of electricity will cease when the potentials have become equal.

Since metals are good conductors of electricity it follows that every charge which is given to them will collect on their outer surface, since the various portions of the charge repel one another, and under the action of these repulsive forces they will move as far apart from one another as is possible, *i.e.* the charges will reside on the surface. These surface charges will distribute themselves so that all points on the surface of the conductor have the same potential, for as long as this is not so the charges will move under the action of the above forces from those parts at a higher potential to those where the potential is lower ; since the body is a conductor, the charges will be able to move under the influence of any electric force which may be present.

For bodies which are non-conductors the above laws, including the one which refers to the motion of electricity, are not valid; non-conductors can have charges at points inside them, and, in general, different points on their surface are not at the same potential.

In order to transfer a quantity of electricity q at a potential V_1 to a point where the potential is V_2 the amount of work necessary is $\pm q(V_2 - V_1)$. The sign depends upon the nature of the electric forces, *i.e.* they are attractive or repulsive according as the charges are respectively unlike or like.

Even when the body is not a sphere electric forces will still be experienced when it is charged, so that work must be done whenever its charge is changed, but we can no longer say that the charge required to raise its potential by unity is equal to the radius. It is still true to say, however, that the charges will cease to move when the potential has become the same at all points and that equation (108) is still applicable.

If a point in the neighbourhood of several conductors is considered, then in order to bring a charge to that point work must be done against all the forces in that field; the total work done is equal to the sum of all the work done against the forces arising from each individual charge, *i.e.* **potential is an additive function** (due regard being paid to the signs). The potential of a body therefore increases when it is brought into the vicinity of a body carrying a like charge and decreases when the charges are unlike.

When there are n charges, $q_1, q_2, q_3 \ldots q_n$, and the distances of some point from them are $r_1, r_2, r_3 \ldots r_n$ respectively, the potential at that point is given by

$$V = \frac{q_1}{r_1} + \frac{q_2}{r_2} + \ldots + \frac{q_n}{r_n} = \Sigma\frac{q}{r} \quad . \quad . \quad . \quad (109)$$

Now inside a conductor the potential is everywhere the same as on its surface, so that no work is necessary in order to move a charge from one point inside a hollow conductor to another similar point. This means that inside a conductor the electric forces are everywhere zero. For a spherical conductor it is easy to show that this must necessarily be so if the charge distributes itself uniformly over the surface of the sphere (on account of the shape of a sphere no other distribution is conceivable), and if Coulomb's law that the force is inversely proportional to the square of the distance is correct. For conducting bodies having any other shape the fact that there is no force inside them can be shown experimentally.

The Potential of the Earth. In determining the work done in moving through an electric field we always set out from some point on the earth and approach the point in question, *i.e.* we always measure the **potential difference** between the earth and this point. If this value is simply called the potential at the point we have tacitly assumed that the potential of the earth is zero; in other words, we have only measured the potential relative to that of the earth. In general, this is the only kind of measurement which we are able to make. For example, when we determine the kinetic energy of a body, we regard the velocity of the earth itself as being zero, and call the difference between the velocity of the body and that of the earth, *the* velocity of the body. The total kinetic energy of a body would be much greater when measured against a body absolutely at rest. But since we are entirely ignorant of its absolute value, and since this additional energy does not appear in any terrestrial phenomena, we should only have always to bear in mind the existence of this fixed amount, if we wished to take it into account. Further, we could not determine the internal energy of a body since the formula on page 198 only refers to the increase or diminution in energy which may be produced. Thus all measurements made on this globe are merely relative, *i.e.* they are based on some convenient but nevertheless arbitrary standard.

Electrical Energy. A further consequence of Coulomb's law is that work must be performed when an electric charge is given to a body, *i.e.* when one wishes to charge the body electrically. If a sphere of radius r_0 has a very small charge Δq, and a second charge Δq is added to it, then the work done is $\Delta q . \frac{\Delta q}{r_0}$. The total charge is then $2\Delta q$. When the next charge Δq is brought up, the work done is $\Delta q . \frac{2\Delta q}{r_0}$, and when the nth charge is brought up so that the total charge on the body becomes $n\Delta q$, then work to be performed is $\Delta q \; (n - 1) \frac{\Delta q}{r_0}$. When we add together all these quantities of work and neglect unity with respect to n (which we can always do, since by diminishing q the number n can be made very large), we have

$$W = \frac{(n\Delta q)^2}{2r_0} = \frac{q^2}{2r_0} \quad [\text{if } n\Delta q = q].$$

The charged body is able to do this amount of work when its charge is again dispersed into small portions at great distances from one another. If we put $q = Vr_0$ we have

$$W = \tfrac{1}{2}\frac{q^2}{r_0} = \tfrac{1}{2}qV = \tfrac{1}{2}V^2 r_0 \; . \qquad . \qquad . \qquad . \quad (110)$$

3. ELECTRIFICATION BY INFLUENCE

The question as to whether all bodies, under all circumstances, can be electrified by a mere redistribution of their charges is of the greatest importance in our attempt to form a true conception of the material world. The following experiment will convince us that this is so. If a long metal cylinder with rounded ends, Fig. 104, is placed near to a positively charged sphere A, no contact being permitted, it can easily be shown that the end C which is nearer A acquires a negative charge whilst the more remote end B becomes positively charged. The process is referred to as **electrification by influence.** The two charges are equal to one another, for if the charged sphere A is removed the body BC is found to be unchanged. The type of charge at B or C can be ascertained by means of a proof plane, which is really an insulated metal disc. It is allowed to touch the body BC, and then removed, a small charge being carried away with it; by allowing a charged electroscope to be acted upon by the charge on the proof plane the nature of this latter charge can be ascertained. In spite of the fact that the body BC has these two different charges, yet the potential is everywhere constant, or, rather, the distribution of charge is such that the potential is the same at all points in the conductor. If we again con-

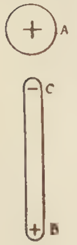


FIG. 104.

sider the long cylindrical body in the neighbourhood of the charged sphere, then, on account of the differences which exist in the distances between the various parts of the body and the sphere, the potential will not be the same initially at these different positions. If the sphere is charged positively, the positive charge in *C* will be repelled, and if the body is a conductor it will move under the action of this force towards *B*, whilst negative electricity will be attracted by the sphere and collect at the nearer end *C*. Now the repulsion which a positive charge at *C* experiences, will be diminished or reduced to zero, because the positive charge collected on the extreme end *B* counteracts the motion, and the negative charge on *C* likewise acts in a contrary direction to the repulsion due to *A*. The final distribution of electric charge is such that the three charges together, namely, the charge on the sphere *A* and the two induced charges at *C* and *B*, produce exactly the same potential at all points on *CB*.

As long as there is a difference in potential between the various parts of the body *CB*, the small elements of electric charge in it experience a force endeavouring to displace them. The direction of the force is such that the motion of the electricity tends to be stopped, and it does actually cease when the forces tending to move an electric charge into some other position have become zero, *i.e.* when all points on the body *CB* are at the same potential. Since the displacement of electricity by influence raises the potential at the more remote end *B* and lowers that at *C*, where the potential was originally higher on account of its closer proximity to *A*, it follows that the final potential of *CB* must be equal to that which initially existed at some point situated near to the middle of *CB*. This point will be permanently without a charge.

If the conductor *CB*, whilst still in the field around *A*, is connected to earth by means of some other conductor it must assume the potential of the earth. Moreover it must lose its positive charge, because its potential was previously greater than zero. When *CB* is again insulated and removed from the presence of *A* it exhibits a negative charge; *i.e.* it has been charged by induction.

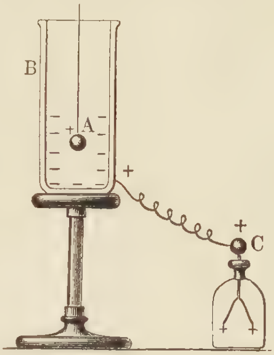


Fig. 105.

Faraday's Ice Pail Experiment. Electrical Screening. Some very *synoptical* and quantitative relationships are obtained when the conductor *CB*, Fig. 105, is so shaped that it completely surrounds *A*. Let us assume that *A* is suspended inside the cylinder *B* by means of an insulating thread (if the cylinder is sufficiently deep, then the disturbing effect produced by the opening which is necessary in order to insert *A* is negligible). Let *B* be connected to an electroscope. When *A* is charged positively and then placed inside *B*, but without coming into contact with it, the leaves of the electroscope are deflected and indicate the potential of *B* at once. It can quite easily be shown that the electroscope is indicating a positive potential.

When the charged sphere is so far lowered that it touches the base of *B* all its charge will pass at once to the cylinder, and, since the cylinder is a conductor, it will pass to its outer surface and there be distributed so that the potential is the same at all points. But however carefully the electroscope is observed, not the slightest indication of any change will be discerned at the instant when the sphere touches the cylinder. It therefore follows that the positively induced charge was exactly as large and similarly distributed before as it is afterwards. Let the charge on the sphere be $+ Q$, whilst $- Q_1$ and $+ Q_1$ are the charges due to influence on the inner and outer surfaces of the cylinder respectively ; these induced charges must be equal in magnitude according to the laws of the conservation of electricity. Since there was no change in the deflection of the electroscope when contact had been made, it follows that $Q = + Q_1$ since the distribution on the outer surface, as experiment shows, is not changed when contact is made. But after making contact the charge Q is certainly on the outside of the cylinder, so hence the charge Q on the outside of the cylinder was previously also equal to Q. Thus all the three charges Q, $- Q_1$ and $+ Q_1$ are numerically equal to one another when the exciting body is completely enclosed.

If the charged sphere is introduced inside but not allowed to touch the cylinder, which is now supposed to be connected to earth so that its potential is zero, then on breaking the earth connexion and removing the charged sphere A without touching the cylinder, the leaves of the electroscope will diverge. A suitable test will show that it has received a negative charge. The magnitude of this deflection is exactly the same as when the charge Q was distributed over the cylinder. The charge is therefore now $- Q$, which is the charge $-Q_1$ originally induced on the interior wall of the cylinder. The charge $+ Q$ was removed when the cylinder was earthed.

The Phenomenon of Electrical Attraction. The process of electrification by influence or induction which has just been described throws new light on the phenomenon of attraction, which is to be observed between charged and uncharged bodies. Initially the uncharged body, say a small piece of paper, acquires two induced charges under the action of a charged body. Now the small piece of paper is generally connected to earth so that charge on it which is similar to that on the exciting body flows to earth. The paper itself retains a charge of the opposite kind so that **attraction then takes place between two unlike charges.** One will remember what was said concerning the law of action and reaction [*cf.* p. 36], and will see how perfectly the law has been ratified here. At first the uncharged body becomes charged, **after which the two electrified bodies attract one another.**

When a charged body is brought near to an uncharged electroscope the leaves of the latter begin to diverge even before any contact has been made. It receives an induced charge of like sign which is excited in the electroscope by the charged body.

Condensers. From statements which have been made previously, an insulated sphere (Fig. 106) of radius r carrying a charge Q has a potential $V = Q/r$; the capacity is therefore equal to the ratio of the charge Q to the potential V, *i.e.* $r = Q/V$. If now a hollow sphere B_2, connected to earth, surrounds the first, then it will acquire a charge $- Q$ on its inner surface, so that if its radius is r_2 its potential if it were alone would be $- Q/r_2$. But now, when both spheres are charged, the potential on B_1, if r_1 is its radius, is

$$V' = Q\left(\frac{1}{r_1} - \frac{1}{r_2}\right)$$

Its capacity, C, calculated as the quotient Q/V', is therefore given by

$$C = \frac{Q}{V'} = \frac{r_1 r_2}{r_2 - r_1} \quad . \quad . \quad . \quad . \quad (111)$$

The proximity of the second and earthed sphere has therefore changed the capacity of the first one; in fact, for all values of r_1 and r_2 it will be increased. The increase will be greater the smaller $r_2 - r_1$ is made, *i.e.* the more closely the second sphere approaches the inner one.

For example, let us assume that $r_1 = 10$ cm. so that its capacity alone is 10 cm.; if it is surrounded by a second hollow sphere of radius $r_2 = 10{\cdot}1$ cm., then the capacity according to equation (111) is 1010 cm., *i.e.* it has been increased a hundred-fold. Hence in order for the inner sphere to be at the same potential as before its charge must be increased one hundred times. It therefore follows that, by bringing an earthed conductor near to an insulated one, a very simple means has been obtained whereby a large quantity of electricity can be stored in a very small space. Such contrivances are called **condensers.**

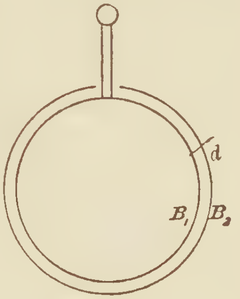


Fig. 106.

Instead of using the radii in the above formula other variables can be introduced. If r_1 and r_2 are nearly equal, one may write more simply r^2 instead of $r_1 r_2$; since $4\pi r^2$ is the area A of the surface of a sphere, we therefore have

$$C = \frac{A}{4\pi d} \quad . \quad . \quad . \quad . \quad . \quad (112)$$

if d is put equal to $(r_2 - r_1)$, *i.e.* d is the distance between the two spheres.

The same meaning can be given to all the variables in the above equation even when the conductors are not spheres. The more exact theory of electrostatics has shown that the capacity of any form of condenser is approximately expressed by the above equation (112). The deviations are due to the fact that some of the lines of force do not pass by the shortest possible path from one plate of the condenser to the other: the deviations are more pronounced for a condenser consisting of two discs than when it comprises two cylinders [*cf.* Fig. 107], and also when the size of the condenser is reduced.

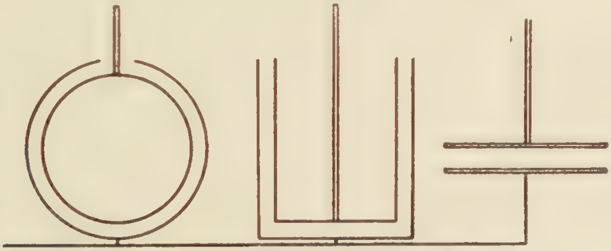


Fig. 107.

4. The Effect of the Medium

All the deductions which have hitherto been made require to be supplemented in a way which **Faraday** first indicated. It was he who enriched our knowledge of electricity by a whole series of new facts; with an almost prophetic insight he suspected them, afterwards subjecting them to experimental confirmation. His achievements have placed him in the front rank of experimental physicists for all time. When his experimental ability was greatest he was engaged, during the years 1837–1839, in studying some problems of electrostatics: these dealt with the phenomena of electrostatic induction which we have already mentioned, the effect of the medium on the phenomena of electrostatic repulsion and attraction, which we are now about to discuss, and other problems covering a much wider range.

If we consider once again equation (102), which merely expresses Coulomb's law, we shall find that it contains the distance of the two charges from one another, exactly as in Newton's law of gravitation where the distance between the two attracting masses occurs. The idea of an unassisted action at a distance between two bodies was something which Faraday could not conceive. Many others had already held such a view but Faraday proceeded to ascertain the influence of the medium by means of experiments. In his hands the subject was advanced further, in a very short time, than it had been through all the meditations of these other people spread over many years. His rare talents were such that he did not waste his time dreaming about insoluble problems, but restricted his work to obtain a solution with the means which were available in those days.

Faraday had no doubt concerning the action of the field, which was the name he applied to the region in which electric forces existed, but he wished to demonstrate it by actual experiment. Even his first ideas on this subject yielded some important results. The space between two charges can only be occupied by insulators, but there are many different insulators which can be used for this purpose. Faraday therefore took two charged spheres and investigated whether the effects which they exert upon one another depended upon the nature of the intervening medium, *i.e.* whether they remained unchanged when one medium was substituted for another. He arrived at the following conclusions.

The forces of attraction and repulsion are influenced to a very great extent by the nature of the medium; for all substances he found that the forces were less than when air was the medium. Later experiments, using more precise and refined methods, have shown that the forces are a little less when air or other gases occupy the space between two charges than when the space is completely free from gas. Faraday characterized all insulating substances by means of a numerical constant which he called the **dielectric constant**; it indicated how many times the effect of any charges were weakened when the space between the charges was occupied by that particular substance instead of by air. The formulæ which we previously obtained are therefore only valid when air is the dielectric medium between the charges. If they are to be applied to other substances, then the value of the dielectric constant, κ, must be inserted; it will have to be inserted in the denominator because the force is reduced when media having greater dielectric constants are used. Coulomb's law is then expressed by the equation

$$f = \frac{qq'}{\kappa r^2} \quad . \quad . \quad . \quad . \quad . \quad (113)$$

The values of the dielectric constants have often been determined exactly and we shall have to refer to their significance on several other occasions. A few values are therefore given in Table 25, opposite.

Since the forces change when the dielectric is altered, it follows that the amount of work done in moving an electric charge through the field

TABLE 25

Dielectric Constants (κ)

Gases.	κ	Liquids.	κ	Solids.	κ
Vacuum	1·0000	Paraffin-oil	2·1	Paraffin wax	2·1
Hydrogen	1·000264	Rapeseed-oil	2·85	Ebonite	2·7
Air	1·00059	Linseed oil	3·35	Shellac	3·1
Methane	1·00094	Ethyl alcohol	27	Sulphur	4·0
Carbon di-oxide	1·00096	Methyl alcohol	35	Mica	6
Ammonia	1·0072	Formic acid	60	Porcelain	4–7
Water vapour	1·0070	Water	80	Glass	5–10

is different, *i.e.* the potential at a point is different. The work done will be κ times smaller since the force is reduced in the ratio $1 : \kappa$. We therefore have

$$V = \frac{q}{\kappa r} \quad . \quad . \quad . \quad . \quad . \quad (114)$$

In order to raise the value of V by unity, the charge must be increased by κr. This means that the quantity of electricity necessary to raise the potential by unity is κ times greater: the capacity has therefore been increased κ times. If we repeat the whole argument regarding condensers as we have done previously, we shall find the factor κ occurring in all the equations. We shall therefore find finally that, instead of equation (112) which represents the capacity of a condenser when air is the dielectric (air-condenser), the capacity of a condenser when the space between the plates is filled with a medium whose dielectric constant is κ, is expressed by

$$C = \frac{A\kappa}{4\pi d} \quad . \quad . \quad . \quad . \quad . \quad (115)$$

These phenomena can also be used in order to ascertain the effect of the medium on the charges of a condenser. It is easily shown that, when the dielectric is charged but the potential difference kept constant, the charges vary in magnitude.

Polarization. Experiments such as the above proved beyond all doubt that the nature of the medium did affect electric phenomena. To-day much more is known about this subject. We know what takes place in a medium when it becomes the seat of electric forces. Let us begin by considering what happens when conductors are placed in an electric field. We have seen that when they are placed in the vicinity of a charged body they acquire a double charge; a like charge at the remote end of the conductor and an unlike charge at the end which is nearer to the charged body. These charges are produced by the electric forces and are made manifest because the charges are free to move in a conductor. In the case of non-conductors the first condition is still true, but the second, referring to the motion of the charges, is not valid. But if insulators do contain electricity just like other bodies the fact that they do not conduct electricity must be because the electricity in insulators is not free to move, but constrained to the various molecules or atoms. It is necessary, however, to imagine that the charges are free to make very small displacements, like the molecules of rigid bodies which are not absolutely immovable but only held together by very great forces. The carriers of these small electric charges would behave like small conductors embedded in an insulating medium. In virtue of the phenomenon of electrostatic influence they receive a charge at both ends, but these charges can only be displaced very slightly. If we have a large number of such small bodies arranged in series one after the other, then the positive charge on one will be very close to the negative charge on its neighbour. The effect of these two charges on outside objects will be practically zero, but the positive charge on the first particle in the series and the negative charge on the final particle are not compensated so that they can be detected. Consequently the following experiment can be made. A piece of sulphur, or paraffin wax, is suspended by means of a well-insulated thread between two charged spheres, Fig. 108. If the spheres are uncharged, then the slab will come to rest in some arbitrary position determined by the residual torsion in the suspension; but if the spheres have unlike charges, then the suspended slab always turns so that it finally comes to rest with its longer axis lying along the line joining the centres of the two spheres. It therefore follows that forces must exist at both ends of the slab, and that these must be forces of attraction so that the charges at the ends of the rod must be opposite to those of the spheres which are nearer to them. The rod is said to be **polarized**, *i.e.* each small constituent part has both a positive and negative pole. But if the rod is broken whilst it is in the field, then new charges will be formed at the two broken surfaces—one on each, just as when a magnetized needle is broken. Taken as a whole, however, each portion will appear to be uncharged just like the original piece of dielectric, because the positive and negative charges are always equal. Now these phenomena can be observed in vacuo where, although they are less in magnitude, their nature is exactly the same as for solid bodies. We have therefore to assume that empty space can be polarized although perhaps in a somewhat different form. The idea that this space should be an absolute void is quite untenable with this hypothesis, for it is not possible for *nothing* to have polarization of any kind. It is therefore necessary to assume that the space between bodies, and even between their molecules, is filled with a material substance capable of becoming polarized. This substance is called the æther. Since we are only able to detect the state of polarization at the extremities of a body, a direct proof of the fact that the æther can be polarized is excluded, for the medium has no boundaries. The problem is just like that which we cited earlier when we were not able to obtain any direct evidence of the effect of the medium on phenomena due to gravitation. In the present one, however, we have really been more fortunate since we have at least been able to demonstrate the effect with other media.

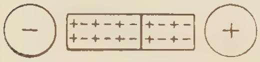


Fig. 108.

Lines of Electrostatic Force. To Faraday the lines of electrostatic force were a graphical representation of the conditions which existed in a field. The direction of the electrostatic force at any point was represented by the direction of the lines of force at that point; the lines of force always had their origin on a positive charge and always terminated on a negative charge. The intensity of the field was indicated by the density of the lines of force, *i.e.* the number per cm.^{2}, whilst the continuity of the lines showed that the force was propagated without any discontinuity from point to point. Although there were originally many physicists who adhered to the idea of action at a distance, yet to-day the theory has been discarded entirely in favour of the more satisfactory conceptions due to Faraday: this has been particularly so since **Maxwell** was able to express these ideas in mathematical form, from which he developed the electromagnetic theory of light.

If the charges on the two spheres, Fig. 108, change their sign, then the polarization in the medium must become reversed, and if the above change of sign of the charges on the spheres takes place continuously the polarization in the medium must reverse with equal rapidity. In Maxwell's theory of electromagnetic waves this reversal of the polarization in the medium plays a very important part.

The nature of electric charges will become more apparent in the following chapters, so that it would be without purpose to continue the discussion at this point.

Conclusion. There is one question, however, which we can answer with greater certainty. In the first place we know that all bodies can be electrified by friction; in this process electricity is not created but merely redistributed. In the second place we know that all bodies can be charged by influence (or induction); under these conditions the two charges which they acquire are opposite in sign, but equal in magnitude. We have therefore to conclude that electricity is quite generally and intimately associated with matter. Since a great force can be called into play even by small quantities of electricity, we might imagine that this force will play an important part in connexion with matter, that the phenomena which occur in the material world will be influenced by these internal electric charges to an extent which even now we can scarcely conceive, and that the influence of these charges will also be extended to atomic phenomena in particular.

CHAPTER III

THE ELECTRIC CURRENT

1. The Laws of Current Electricity

SINCE electricity cannot be created, it follows that a body can only be charged if electricity moves from one place to another. In the phenomena which we have so far discussed, we have always been concerned with the motion of electric charges, but we have not discussed the effects of this motion. We have now to make inquiries concerning the laws which can be applied to electricity in motion.

Introductory Experiments. Let us suppose that we have a large Leyden jar l which is charged (Fig. 109). An electroscope E serves to indicate the potential of the jar. The knob of the jar is connected by means of a piece of string, about 1 mm. in diameter and 1 metre long, to a universal electroscope U [*cf.* p. 234]. This electroscope will discharge itself every time the deflection of its small ribbon-like leaf is so great that it comes into contact with the earthed case of the instrument. In order to do this a definite potential must have been given to the electroscope, and here, according to equation (108), a definite quantity of electricity also.

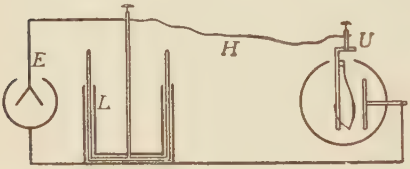


Fig. 109.

If the capacity of the Leyden jar is so great that its potential does not vary appreciably in the course of these experiments, then it will be found that the time which elapses between two successive discharges of the electroscope is constant. When the length of the connecting thread is increased 2, 3, 4, or more times, this interval will be increased in the same ratio. This fact indicates that it requires this extra time in order for the same quantity of electricity to pass through the cord and raise the potential of the electroscope to the same value, or, the quantity of electricity passing per second is less when the cord becomes longer.

If we then place a second cord identical to the first alongside it the electricity can flow along two paths from the jar to the electroscope. What we expect is found to occur: the electricity uses the two paths and the interval of time between successive discharges is reduced to one half its former value. With 3, 4, 5 or more equal cords similarly placed this interval is reduced in the same ratio.

Let us now suppose that these several cords are joined together to form a thick rope; will anything else be changed? Perhaps this question may be answered negatively at once since the path which is at the disposal of the electricity has not been changed. Experiment shows that this suggestion is correct. From the point of view, however, in which logical proof is required, we have to reply that this is by no means self-evident. For if we assume again, as we did when discussing the seat of electrical charges, that the charge adjusts itself by passing over the surface of the conductor, we should have to consider the fact that the surface area of the cords is diminished when they are twisted together, for the greater part of the surfaces of the individual strands would now be placed inside the rope. Since the experiment indicates no change in the time which we are discussing it follows that the whole cross section must be available for the passage of electricity.

When we use threads of the same diameter and length but made of different materials, or if we change the condition of the same thread by moistening it, we shall find that the time of discharge is different on each occasion. We say that different substances do not conduct electricity equally well.

If we try to investigate metal wires in this way we shall soon find that it is impossible. The very first time that the electroscope was deflected a spark would be seen when contact took place, the leaf of the instrument would be burnt out, and the Leyden jar completely discharged. The metal wire is such a good conductor that during the short time of contact the whole jar becomes discharged.

Let us now return to our first experiment, but, whilst using the same cord, let the potential of the jar be reduced to half its former value ; the time of discharge will be doubled. When the potential is reduced still further the time of discharge will increase accordingly. This diminution in potential can be obtained by connecting the jar to an equally large but uncharged one for a short time. The charge Q on the first flash is distributed between the two jars, *i.e.* the capacity C' on which the charge now resides has been increased twice. The potential must therefore be one half its former value, for according to equation (108) we have $Q = CV = C'V'$, so that $C : C' = V' : V$. Since $C' = 2C$, it follows that $V' = \frac{1}{2}V$.

Ohm's Law. Let us now collect the above facts into one single and comprehensive law. Since from the period between two discharges we make conclusions concerning the quantity of electricity which has passed along the thread per second, *i.e.* the current I, we have for our definition of **current** the expression

$$I = \frac{Q}{t} \qquad . \quad . \quad . \quad . \quad . \quad (116)$$

From the above experiments we are therefore able to formulate two laws :

1. The current is greater when the potential difference applied to the two ends of the conductor through which the current flows is increased.

2. Moreover the current also depends upon the nature of the conductor, so that to every body we ascribe a definite conductivity for electricity S. The greater the current the greater is the conductivity of the material. Whilst at first we again measure this conductivity in arbitrary units, we are able to say that the quantity of electricity which passes per second, *i.e.* the current, is proportional to the potential difference V and to the conductivity S. In symbols we have

$$I = kVS,$$

where k is a constant ; in this equation we assume that the potential V and the conductivity S remain constant.

We will now endeavour to make the constant of the above equation equal to unity, as we have done before, by a choice of suitable units. In the present instance the unit of current has already been defined in terms of the units of quantity and time, and the unit of potential has likewise been expressed in ergs so that it is only possible for us to eliminate the constant by a suitable choice of the unit used to measure the conductivity. A conductor therefore has unit conductivity when a unit potential difference causes a unit current to pass through it ; then

$$I = VS \qquad . \quad . \quad . \quad . \quad . \quad (117)$$

Further, we have already seen that the conductivity of a body depends upon its dimensions. The longer the circuit, the less is the conductivity, whilst if the cross section is increased the conductivity becomes greater ; moreover the nature of the material also helps to determine the conductivity. In order to be able to compare the conductivities of different substances it is necessary to use pieces of the same length and cross-sectional area for this purpose. We therefore speak of the **specific conductivity** or **conductance** of a substance, a term which is used to indicate the conductivity of a cylinder 1 cm. long and having a cross-sectional area of 1 cm^2. If we denote this by σ, whilst l and a are its length and cross-sectional area respectively, we have, according to the above,

$$S = \frac{\sigma a}{l} \qquad . \quad . \quad . \quad . \quad . \quad (118)$$

A body will also have a unit conductance when it is in the form of a cube whose edges are 1 cm. long.

The law which has been expressed by equation (115) can also be expressed by

$$I = \frac{\sigma a}{l} V \qquad . \quad . \quad . \quad . \quad . \quad (119)$$

Resistance. The above phenomena can also be considered from quite a different viewpoint. The potential difference tends to distribute the electricity, but the thread determines how quickly this shall take place ; it allows a certain quantity to pass per second, and no more. Instead of saying that the thread possesses a certain conductivity, we can also say the thread offers a certain **resistance** to the passage of electricity. Experiment reveals the fact that the resistance is greater, the longer the thread and the smaller its cross section.

The current is directly proportional to the potential and inversely proportional to the resistance. If the resistance is R, we have

$$I = k\frac{V}{R}.$$

If the unit of resistance is now defined as the resistance of a conductor which is such that unit current passes when the applied potential difference is unity, $k = 1$, and we have

$$I = \frac{V}{R} \qquad . \quad . \quad . \quad . \quad . \quad (120)$$

and if the **specific resistance or resistivity** ϱ is defined in the same way as the specific conductivity we have

$$R = \frac{\varrho l}{a} \qquad . \quad . \quad . \quad . \quad . \quad (121)$$

The resistance of a body is always inversely proportional to its conductivity. The one or other manner of expression may be the more convenient, but there is no essential reason why one should be used exclusively in preference to the other.

The mode of treatment here used was chosen to preserve the continuity of ideas in our discussion of the subject. The above experimental arrangement is not suited to verify the law exactly, the main reason for this being that the potential difference across the two ends of the string immediately after the discharge of the electroscope is greater than it was just before.

We shall soon become acquainted with other means whereby the current can be measured more exactly, and which are also suitable for the verification of the law in the case of metallic conductors.

The law which is expressed by the equation $I = V/R$ is called **Ohm's law** (1826), and it is the fundamental law of current electricity. It is very frequently used in electrotechnology.

Joule's Law. According to the remarks made on page 235, the passage of an amount of electricity Q from a potential V_1 to a lower potential V_2 is associated with a loss in potential energy equal to $Q(V_1 - V_2)$. If instead of $V_1 - V_2$ we write V for simplicity, and since $Q = It$, it follows that the loss in energy in time t is $W = IVt$. Now according to Ohm's law $V = IR$, so that we may write for this loss in energy $W = I^2 Rt$. When a current I passes along a conductor whose resistance is R, this energy is converted entirely into heat, so that we have the following law: the amount of heat developed per second when a current I is flowing in a resistance R is

$$W = I^2R \quad . \quad . \quad . \quad . \quad . \quad (122)$$

This law was formulated by Joule in 1841. As expressed above it is only true when all the variables are measured in absolute units [*cf.* p. 262].

The amount of heat developed is therefore greater when the resistance is increased. In this statement, however, one sees a certain inconsistency, for when the resistance of a conductor is increased the current through it is less. What the above formula really asserts is that, if the current remains constant, the heat developed is proportional to the resistance of the conductor. In order to pass the same current through a wire of higher resistance it is necessary to increase the potential difference between its two ends, and the fact that more heat should be developed by the same current when the potential is increased no longer causes surprise. The law which is expressed by equation (122) is called **Joule's Law,** and it is the second fundamental law of electrotechnology. It is this heat which raises the filaments in electric bulbs to incandescence; these lamps are generally arranged so that the same potential difference is applied to each one; it is therefore more convenient to express Joule's law in terms of V and R. Since $I = \frac{V}{R}$, we have $W = V^2/R$, *i.e.* the lamps which have the smaller resistance develop most heat. From our definition of potential, the energy developed is expressed in ergs, and the energy per second in ergs/sec. If the heat developed is required in calories, then we must remember that $4{\cdot}2 \times 10^7$ ergs or 4·2 Joules are equal to one gram-calorie.

The heat which is developed by a current can be applied to measure currents more accurately as in the so-called hot-wire instruments (Fig. 110). The current is passed through a wire AB having a high resistance; this wire is heated by the current and expands accordingly. This increase in length is used to move a pointer across a calibrated scale, so that the strength of the current can be observed at once.

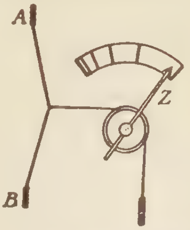


Fig. 110.

Closed Circuits. Whenever electricity is in motion, the motion requires the existence of some cause; it is assumed to be the electrical force due to a potential gradient. Such differences of potential arise whenever electricity accumulates at some point, *e.g.* on the coating of a Leyden jar. The electric current tends to equalize the potential so that the Leyden jar loses its charge, its potential falls, and when this has become equal to that of the earth the current ceases. It is exactly the same with a stream of water. For water to flow it is necessary to have two containers, one at a higher level than the other; the lower container may be the sea, or a low-lying reservoir. The current causes the level of the water in the higher container to fall whilst the lower level rises, until when both are equal the current ceases to flow. Now all vessels capable of storing electric charges, even very large condensers, would be completely discharged in a very short time, amounting in fact to no more than a mere fraction of a second, if they had to supply a current sufficient to accelerate a train or drive a machine. If the current is to last for a considerable time, then the source of electricity must be replenished, and if the current is to be regular, then the energy supplied to it per second must be equal to the energy which is dissipated. This means that we need a source which will supply energy continuously. If we still retain the analogy of the water, not only must the water in the higher reservoir be replenished, but the lower one must be emptied continuously; in other words, we need a force which will cause the water to return again from the lower reservoir to the one above.

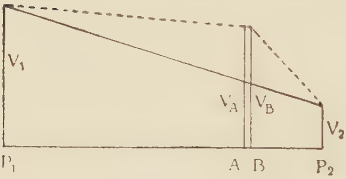


FIG. 111.

Hence it follows that a continuous supply of water can only be obtained in a **closed circuit** in which there is a machine at some point which does work and raises the water again to the higher reservoir. This process is seen in miniature in a hot-water system, where the boiler not only serves to heat the water but also to raise this water which has become less dense. This same process occurs on a larger scale in Nature. Water from mountain springs and lakes flows through pipes to the turbines from whence it passes into the rivers and finally to the sea. The sun pours forth energy upon the sea continuously, causing the water from extensive areas to evaporate and rise into the air, from which it is precipitated again as rain or snow. Unless the sun did raise the water and so complete the cycle all water-wheels would soon come to a standstill for lack of water.

The conditions are somewhat similar when we have a continuous flow of current; there must always be a closed circuit in which is included some contrivance for supplying the electrical energy. Small pieces of apparatus such as electric bells are operated by means of a dry battery, whilst powerful dynamos supply the energy in a large electrification scheme.

Now in all such processes as these where we have a continuous current, the fundamental law which holds is that the quantity of electricity passing any cross-section per unit time is the same. Let P_1 and P_2, Fig. 111, be the two poles of some battery, or even dynamo, whilst V_1 and V_2 are their potentials respectively. The electricity will then move along the conductor from P_1 to P_2, the potential falling from its highest value at P_1 to its lowest at P_2. Now let A and B be any two quite arbitrary points in the circuit. If then more electricity flows into A than flows out through B it follows that the piece of conductor AB will become charged after a short time. Its potential would therefore rise, *i.e.* its value would approach that of V_1, or the difference $V_1 - V_A$ would become less, whilst $V_B - V_2$ would similarly increase. The necessary consequence of this would be that, according to Ohm's law, the current between P_1 and A would be diminished whilst that from B to P_2 would increase. This would continue until the current arriving at A was equal to that which flows from B. This same theory may be applied to water flowing along a closed set of pipes.

2. THE THEORY OF METALLIC CONDUCTION

All the statements which have hitherto been made concerning electricity in motion have only been of a phenomenological nature, *i.e.* they merely describe the motion without allowing us to obtain any insight into the phenomena. If we now proceed to discuss those attempts which have aimed at giving us an explanation, we have to acknowledge at once that the process of metallic conduction has proved extraordinarily difficult to account for satisfactorily.

The first important point which presented itself in this connexion was the fact that the change of electrical resistance with temperature is very nearly the same as the coefficient of pressure increase for a gas (**Clausius**, 1858). It will be remembered that this coefficient for a perfect gas is 0·00366, and it has been found that the resistance of metals increases by nearly 0·00366 of its value per unit rise in temperature. The mean values over the range of temperature 0° – 100° C. for a few metals are given in the following table.

TABLE 26

Metal.	Mean Coefficient of Increase of Resistance.	Metal.	Mean Coefficient of Increase of Resistance.
Aluminium	0·0042	Manganese	0·0038
Lead	0·0041	Platinum	0·00367
Gold	0·00367	Silver	0·0040
Copper	0·0037	Cadmium	0·0037

These facts prove conclusively that there is a connexion between electric conductivity and the gaseous state.

Moreover, it was very soon noticed that the very best conductors of electricity are those metals which have a high thermal conductivity. Silver is the best conductor for both heat and electricity (Law of **Wiedemann** and **Franz**). The connexion between the two properties is unmistakable, even though the law does not possess absolute validity. It was even found that both increase with the absolute temperature in almost the same proportion.

But the early attempts to give a theoretical explanation of the nature of these two processes were not successful ; the failure was mainly due to the fact that nobody at that time had any idea (or only a very imperfect idea) concerning the nature of electricity. We are indebted to **Riecke** (1845–1915) for the first serious experiments made in this direction. He allowed a very large current to pass for an hour between two carefully polished cylinders of copper and aluminium in order to see whether or not any material particles were carried forward by the current. He could not observe the least trace of any metal which had been transferred from one cylinder to the other, so that he was forced to conclude that the electric current did not carry any particles of matter with it.

A further advance concerning this question was made when it was discovered that electricity consists of very small negatively charged particles, which were not associated with any form of gravitational mass then known. **Sir J. J. Thomson** assumed that such negatively charged particles, called **electrons**, are free to move in the interstices between metallic atoms just like the molecules of a gas. From this idea **Drude** obtained an explanation of the law of Wiedemann and Franz. The metal particles are vibrating and the free electrons take up some of this thermal energy, and so acquire the ability to conduct heat. Now whereas the atoms cannot move from one place to another in the solid body, the electrons are able to pass with great speeds throughout the entire body. If at one point in the body a higher temperature prevails than elsewhere, then those electrons which come from this place will have velocities which are greater than those of other electrons. These electrons with a higher speed will collide with atoms in cooler parts of the body, whereas those electrons which come from the colder regions in the body will travel with smaller velocities to the hotter parts. There they will acquire a greater velocity and return to parts of the body at a lower temperature.

If now a potential difference is applied between two points in a metal, then it superposes on the otherwise haphazard motion of the electrons a tendency for the electrons to move in one direction ; this direction will be opposite to that of the applied force since the particles are negatively charged. When electrons collide with the atoms of a metal they lose their energy, but at the same time the irregular vibratory motion of the metallic atoms becomes more vigorous. Electrical energy will therefore be converted into thermal energy according to Joule's law. But since the electrons which, moving against the direction of the applied electromotive force (*E.M.F.*) have a higher velocity than those moving in the reverse direction, it follows that there will be a preponderance of electricity conveyed in one direction ; **this conveyance of electricity constitutes the electric current.**

In different metals the number of free electrons per cm.^{3} varies, or, in the language of the Kinetic Theory, we may say that the electron-density, and therefore the mean free-paths, vary. This will account for the differences which exist in the thermal and electrical conductivities of different metals.

The assumption that it is only the electrons which convey the current in metals has been confirmed in recent years by **Tolman's** experiments. He has shown that the carriers of the electric current in metallic circuits have all the known properties of electrons. The ratio of the electric charge to its mass, viz. e/m, is the same as for electrons. In developing this theory, however, several difficulties have been encountered. In the first place, if electrons do convey thermal energy, then another degree of freedom will have to be added to those already possessed by metallic atoms. This will cause an increase in the value of their specific heats by an amount depending upon the number of degrees of freedom already existing [*cf.* p. 156]. **Pauli** has met this difficulty by assuming that the extremely light electronic gas behaves at ordinary temperatures as gases do in the neighbourhood of the absolute zero, where, according to Nernst's Third Law of Thermodynamics, the specific heat becomes zero, or at least practically zero.

Since the year 1920, **Kamerlingh-Onnes** (1853–1926) has shown that the resistance of several metals at the low temperature of liquid helium suddenly becomes so extraordinarily small that it cannot be measured. He refers to this as the superconductivity of metals. These phenomena have been observed with mercury, thallium, lead, tin, and indium, at temperatures below 4·25, 1·5, 2·45, 1·8 and 3·4 degrees on the absolute scale respectively. It was therefore possible to pass a current of 1200 ampères through a column of mercury 1 mm.^{2} in section without heating it appreciably. A magnet was used to induce a current in a lead ring [*cf.* p. 264], this current being usually reduced to zero in a very small fraction of a second ; at these low temperatures, however, the current had not been reduced by one per cent. even after one hour. In 1925 **Kamerlingh-Onnes** reported that at these low temperatures the conductivity of a lead wire was 10^{12} times greater than at 0° *C*. If electrons are to be regarded as a gas, then one would expect that they would lose their velocity completely at the absolute zero ; actually it is just the opposite, for they then appear to be able to move more quickly than ever.

In conclusion it may be said that the insight into the nature of electricity which has been obtained during the last few decades has prepared the way for the solution of this problem. Especial mention must be made of **Sommerfeld's** attempts (1927) to explain the conductivity of metals according to ideas which have been developed in the new Quantum Mechanics ; they appear to be clearing the way for a satisfactory solution to these problems.

3. THERMOELECTRICITY

When a piece of iron wire and one of copper are soldered together at both ends so as to form a closed circuit, an electric current is always present in this circuit, so long as the temperatures of the two junctions are different. Such a pair of wires constitutes a thermocouple, and the potential difference which causes the electricity to move according to Ohm's law in such a circuit is called the **thermo-electromotive force** of the two metals. All metals can be arranged in a series, such that the thermo-electromotive force of any pair becomes greater, for the same temperature difference, the further the two metals are separated from one another in this series. Some of the more important metals can be arranged as follows : antimony, iron, cadmium, zinc, gold, silver, copper, aluminium, platinum, nickel, constantan, bismuth. The potential difference for the two extreme metals antimony and bismuth amounts to 0·012 volt, when one junction is at 0° and the other at 100° *C*. If several such couples are arranged in series greater potential differences can be obtained. The strength of the current in such a circuit depends upon the resistance of the whole circuit as expressed by Ohm's law. The potential difference is not proportional to the temperature interval between the two junctions. In practice suitable experiments must be made in order to discover how this *e.m.f.* varies with temperature.

The phenomena of thermoelectricity can be satisfactorily explained in terms of the hypothesis which has been detailed just previously. Let us assume that two strips of different metals are joined together at one end only (Fig. 112). Since, in general, the electronic density, and therefore the electronic pressure, which will be due to the repulsion of equally charged negative particles, will be different in each of the two metals, a process of compensation will take place at the soldered joint. From the body having the higher electronic density (let us say A) some electrons will pass over to B. Since both metals were initially uncharged, A must become positively charged, since it has lost electrons, whilst B which receives them will become negative. This, however, will tend to check the further transfer of electrons since these charges will attract the electrons to A and repel them from B. As soon, therefore, as the potential difference at the junction has become so great that there is equilibrium between the two electronic pressures, no further transfer of electricity can take place. But there will always exist a certain potential difference at this joint. This fact does not contradict the assertion which we made earlier on, that the potential in a metallic conductor must be the same at all points in it, for, in this case, there are forces present which redistribute the charges so that a potential difference does exist. In the previous example we had only shown what would happen if the electrons are only acted upon by forces represented by Coulomb's law.

Since the electronic density, like that of a gas, is proportional to the absolute temperature, so that it is equal to aT for one body and bT for another, where a and b are constants, the potential difference at a junction of two metals must be $(a - b)T$, which is again proportional to the absolute temperature. When therefore the other two ends of the wire are joined together as in Fig. 113, it follows that electrons will tend to pass from A to B at this junction D. The forces at the two junctions will act in opposite directions, endeavouring to send the electrons in opposite ways around the circuit. But since the driving force at the junction D is less than that at C, it can only diminish the effect of the force at C but never reduce its effect to zero entirely. There is therefore a residual $E.M.F.$ in the circuit and this will increase, in general, when the temperature difference between the two junctions becomes greater. In consequence of this $E.M.F.$ an electric current will be produced in the circuit and this will tend to make the potential difference vanish. Since, however, there exists at the junctions some process which causes the potential difference to be re-established, the current of electricity will be maintained as long as the temperatures of the two junctions remain different.

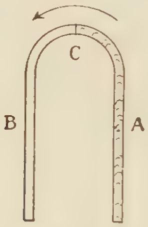


Fig. 112.

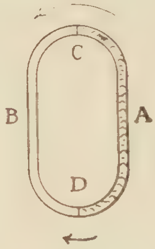


Fig. 113.

A question may now be raised concerning the origin of the energy which produces this current. The answer to it is found in the fact that in the whole process electrons are conveyed from a hot place C to a cold one D, thereby transmitting energy from one point to one at a lower temperature. Subsequently they return across D from the cold junction to one which is hotter. Since they return cold, they will be heated at the warm junction, thereby gaining energy. The thermocouple is therefore a machine in which heat energy is converted into electrical energy.

The Peltier Effect. In concluding this subject another phenomenon known by the name of the **Peltier effect** must be briefly mentioned. If an electric current is passed through a pair of metals, as in Fig. 114, by joining their two free ends to a small battery, something will be observed to occur at the point where the two metals are soldered together. If the direction of the current is such that the thermal *E.M.F.* alone would drive the electrons in the same direction, it would now assist the transfer of electricity, thereby doing work and cool the junction a little. When the current flows in the opposite direction the junction is heated to an extent which is more than can be accounted for by Joule's law. The cooling effect which takes place when the current flows in one direction, and which becomes a heating effect when the current is reversed, was discovered by **Peltier** in 1834.

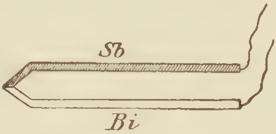


FIG. 114.

CHAPTER IV

ELECTROMAGNETICS

1. Magnetic Effects due to a Current

WHEN, in 1820, **Oersted** was investigating the properties of electric currents in wires near to a suspended magnet, he noticed that the magnet was deflected every time the current was allowed to pass, and that it returned to its original position as soon as the current was broken. **Ampère** (1775–1836) then specified the direction in which the magnet would rotate. If the magnet is suspended so that when at rest it lies in the magnetic meridian, *i.e.* its direction coincides with the earth's horizontal field, and a current is allowed to pass through a wire which points in the same direction as the needle, then the deflection is given by the following rule.

Hold the thumb of the right hand at right angles to the fingers. Place the hand on the wire with its palm facing the needle and the fingers pointing in the direction of the current. The thumb will then indicate the direction in which the north pole will be deflected.

If this rule is applied to various points around a straight wire carrying a current it will be found that the magnetic lines of force are circles with their centres in the wire and their planes at right angles to it.

If, as with magnets, we now wish to represent the direction of the force, *i.e.* the direction in which a north-seeking pole would move, by means of line of force, then complete circles must be drawn so that their centres lie in the wire. If a north magnetic pole is carried around in the other direction work must be done against the forces due to the current, whereas energy is gained if the pole is moved in the former direction, since the forces which already exist in the system will now do the work. We can no longer speak of a potential in the field, however, in the same sense as we have done before. When the pole is brought nearer to the conductor, no work has to be done, since there are no forces present which tend to oppose this motion, but, on the other hand, work must be done when the pole is moved around the conductor. If the pole is moved completely around the circuit the same state of affairs will again be realized when the motion has been completed, if we assume that the current has remained unchanged, except for the fact that the potential at any point in the field has no definite and constant value, but depends upon how often a unit pole has been moved around the circuit.

A picture of the lines of force around a conductor carrying a current can therefore be obtained, as we have already done in the case of magnets, by scattering iron filings on a piece of cardboard which is intersected at right angles by the wire (Fig. 115).

If the wire is bent into the form of a loop (Fig. 116) and one traces, according to the above rule, the direction in which a magnet assumed to be suspended at the centre of the loop, will be deflected by the current at each part of the circuit, it will be found that this direction is always the same. Each part of the circuit gives rise to lines of force which proceed in the same direction from some point inside the loop; these lines encircle the wire and return to form a continuous curve. Now since lines of force always start from a north pole and terminate on a south pole, it follows that the current in a loop gives rise to a north-seeking polarity on one side and a south-seeking polarity on the other. If we imagine a steel disc whose periphery coincides with the wire, and which is so magnetized that one face has north-seeking magnetism and the other south, then the paths of the lines of force outside the disc are exactly like those due to a current in a loop; if it were possible for some person to see these magnet effects he would be unable to say whether they were due to a magnet or a current in a loop.

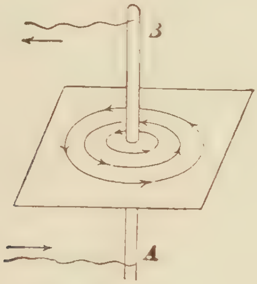


Fig. 115.

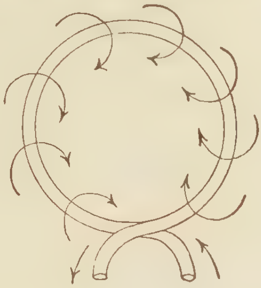


Fig. 116.

When several such wire loops are placed in series, a procedure which is most easily accomplished by winding one unbroken wire to form a spiral, then the current in each complete turn of the spiral forms such a thin magnetic shell, and the effect of the current in the whole spiral (solenoid) is the same as if a series of magnetized steel discs were placed to form a pile. This, however, is exactly the way in which we have conceived a magnet to be constituted—one small magnet is added to another so that like poles all point in the same direction. Hence, when a current is flowing in a solenoid the effect is that of a complete magnet, Fig. 117. At one end the lines of force emerge from the coil whilst they return at the other end. Between any two wires the magnetic effect is zero, since the poles of the corresponding magnetic shells are dissimilar.

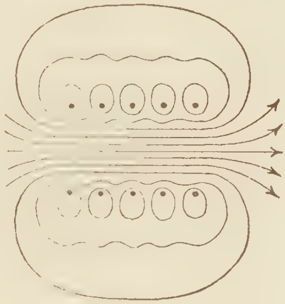


Fig. 117.

In 1879 the discovery of Oersted was supplemented by an experiment due to **Rowland.** He proved that when an electric charge was rotated mechanically the deflection of a magnetic needle was exactly the same as if the same quantity of electricity had passed per second in a stationary circuit.

According to the law that all effects in the inanimate world are due to the action of some body, a twofold conclusion must be drawn from this phenomenon.

1. It is inconceivable that only the magnet is moved by the current, for the magnet must exert some influence upon the circuit; this effect will be equal in magnitude to the first, but opposite in direction. If the

circuit is free to move and is brought near to a magnet, the above force is easily discerned; only a few days after Oersted's discovery Ampère devised an apparatus similar to that shown in Fig. 118. It consists of a wire frame the ends of which dip into mercury cups A and B so that it is free to rotate. By attaching leads from a battery to the mercury cups a current can be passed through the frame. When the current is sufficiently strong, or the windings sufficiently numerous, this frame begins to rotate under the action of the small field due to the earth's magnetism, until its axis coincides with the direction of the earth's field.

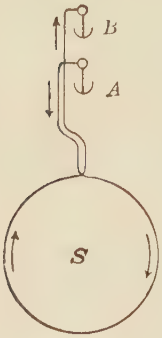


Fig. 118.

2. The second deduction is even more important. If it is correct that only forces of similar nature can act upon one another we have to conclude from the above phenomena that an electric current and a magnet are identical. But are we to consider the current in the conductor as a magnet, or *vice versa*? Now since magnetism is something whose secrets are still hid from us, more than those of electricity, it is more convenient to consider a magnet as the product of a current. This idea was adopted as soon as Oersted published his discovery. Ampère assumed that magnetic effects are due to very tiny circular currents which flow continuously. Let us imagine that the atoms are encircled by small currents. In the unmagnetized state these circular currents will be arranged irregularly. The magnetic condition arises not when small magnets, but when these small circular currents are rotated so that their planes are all parallel. Let us further imagine that we have a large number of atoms each encircled by a current, as in Fig. 119. In the inner circuits there are always two currents very close to one another so that the external effect of one circuit is neutralized by that of another. Hence it is only the circuits on the periphery of the magnet which contribute to the external effect.

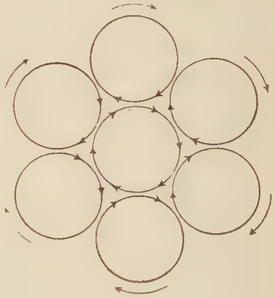


Fig. 119.

The assumption that such continuously circulating currents do exist has one very great difficulty; it is due to the fact that all currents which we have considered in other connexions have to overcome resistance, so that either some electric force is necessary to maintain them or else they cease to flow after a very short time. Whenever they flow, a continuous supply of energy must be forthcoming. This cannot be so in the case of a steel magnet. Nevertheless all recent attempts to explain magnetism are based on the assumption of such molecular or atomic currents. It is held that the above-mentioned reasons are not plausible, since one is here dealing with currents in very tiny orbits, for which the laws applicable to currents in macroscopic circuits are not valid, especially since it has been observed that at low temperatures, when a current has been excited, it will continue to flow for one hour without being reduced in strength [*cf.* p. 253]. Finally, **Einstein** and **De Haas** have detected experimentally a kind of gyratory motion in magnets which can only be explained on the assumption that these small circular currents do exist.

The Law of Biot and Savart. In the same year (1820) **Biot** (1774–1862) and **Savart** (1791–1841) proposed a law which is a numerical expression of the effect between currents and magnets. It is obvious that their law was formed on the model of the laws of Coulomb and Newton. They stated that a small element ds of a conductor carrying a current will exert upon a magnetic pole P a force which, in the first place, is proportional to the current I in the conductor, and the pole strength m. Secondly, the force is inversely proportional to the square of the distance r; thirdly, the force will be greater when the length of ds is increased; whilst, finally, we have to ask whether the force will be independent of the angle θ, Fig. 120, the angle between the conductor ds and the radius vector r. They assumed that the effect will be greatest when the magnetic pole is placed on a line running perpendicular to the element, *i.e.* when it is at A. When the pole is in other positions the effect must be multiplied by $\sin\theta$. The law can be expressed thus:

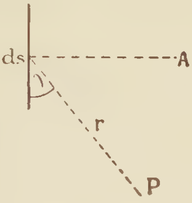


Fig. 120.

$$f = C\frac{Im \sin\theta ds}{r^2} \quad . \quad . \quad . \quad . \quad (123)$$

where C is a constant.

Expressed in this form the law cannot be verified since it is impossible to produce such a short conductor. But starting by assuming that the law is correct, one can calculate the total intensity arising from the current in a complete circuit, *e.g.* the intensity due to either an infinitely long straight conductor, or one which is circular. The formulæ then obtained can be subjected to experimental test. This has been done many times both directly and indirectly, and on no occasion has any evidence been obtained which would contradict the law.

The simplest case to which the law can be applied is that of a circular current when there is a magnetic pole at its centre. The line joining the element to the pole is then perpendicular to the direction of the element, *i.e.* $\theta = 90°$; we must also remember that $\sin\theta = \sin 90° = 1$. Now all the elements of a circular current are at the same distance from the pole, so that the total magnetic force is

$$f = \frac{CIm}{r^2}\Sigma ds,$$

where the summation is to extend all along the circumference of the circle. The result of the summation is $2\pi r$, so that

$$f = CIm\frac{2\pi r}{r^2} = \frac{2\pi CIm}{r} \quad . \quad . \quad . \quad . \quad (124)$$

This means that the magnetic intensity at the centre of a circular current is proportional to the current, and inversely proportional to the radius r. If we do not wish to know the effect due to all the current but only that arising from a length l, we have only to write l instead of $2\pi r$ in the above, so that the force, in this case, becomes $f = CIml/r^2$. When we require the force due to 1 cm. of a circular conductor we put $l = 1$; under these circumstances

$$f = \frac{CIm}{r^2} \quad . \quad . \quad . \quad . \quad . \quad (125)$$

Since all actions must always take place between two participating bodies, it follows that a magnet of pole strength m will act on a current in a circuit with a force equal to the above. If we therefore put $\frac{m}{r^2} = H$, where H is the strength of the external magnetic field at a point in the conductor, and is the same for all points in our problem, we have

$$f = CIH \quad . \quad . \quad . \quad . \quad . \quad (126)$$

This expresses the force or unit length of the conductor-carrying current I, when placed in a magnetic field H.

Moving Electric Charges in a Magnetic Field. On account of the importance which this subject assumes at a later stage, we shall here calculate the force when the current is not continuous but consists of a succession of very small charges separated from one another like pearls on a string. Let us assume that there are n electrified particles on 1 cm. length of the string, each carrying a positive charge e. Let the velocity with which the small particles move be v. Then nv particles pass through any cross-section of the wire per second and carry a total charge nve with them ; this is the strength of the current I. Consequently, according to equation (126), the force on unit length is $CIH = CnveH$. This force is operative upon 1 cm. or n particles, so that the effect on each particle is $f = CveH$. We will anticipate that e will be measured later on in such units that C is unity, so that

$$f = veH \quad . \quad . \quad . \quad . \quad . \quad (127)$$

This equation expresses the force on a charge e moving with a velocity v in a field whose intensity is H. The direction of the force is perpendicular to the line of motion of the charge and perpendicular to the lines of magnetic force. This fact can easily be remembered by the aid of **Fleming's left-hand rule** : If the index finger of the left hand points in the direction of the lines of force of the field, and the middle finger in the direction of the current (*i.e.* the moving charge), then the force on the conductor (charge) will be in the direction of the outstretched thumb.

Magnetic Intensity due to a Coil of Several Turns. Equation (124) enables us to specify very easily the intensity of the magnetic field at the centre of a coil consisting of N turns. Since the term intensity of magnetic field implies the force on unit pole, we have at once

$$f = \frac{2\pi CIN}{r} \quad . \quad . \quad . \quad . \quad (128)$$

For this formula to be valid all the turns must lie so close together that one can assume that the magnetic pole lies at the centre of each turn. This action between magnets and currents is responsible for an immense number of industrial applications ; in fact nearly the whole field of applied electricity depends upon this effect. We shall only mention telegraphy, telephony, and the electric motor.

Galvanometers. The laws which are embodied in equations (124) and (128) serve to measure currents exactly. If a small magnetized needle is suspended at the centre of a circular coil, it will come to rest with its axis in the direction of the earth's field when the coil is not carrying a current, or rather, it will lie in the direction of the earth's horizontal component NS, *i.e.* in the meridian NS. Suppose now that the coil is rotated so that its plane also lies in the magnetic meridian. If a current is passed through the coil, it will endeavour to rotate the magnet into a direction perpendicular to that of the coil, whilst the effect of the earth's field will be to tend to restore it to its original position. These two forces have a resultant which acts on the one pole of the magnet with a force f. At the other pole an equal force will be acting in the opposite direction. The magnetic needle, Fig. 121, which is suspended freely, will finally come to rest in this direction. If H is the horizontal component of the earth's field, the force on the pole m is mH, whilst, according to equation (128), the force due to the current is $2\pi CINm/r$. Consequently the equilibrium position in which the needle finally comes to rest is given by the angle θ, where

$$\tan\theta = \frac{\dfrac{2\pi CImN}{r}}{mH} = \frac{2\pi CIN}{rH} \quad . \quad . \quad . \quad (129)$$

If we again put $C = 1$ the current is given by the equation

$$I = \frac{rH}{2\pi N}\tan\theta \quad . \quad . \quad . \quad . \quad (130)$$

The pole strength vanishes from this equation, since it only affects the two forces in the same ratio, so that this arrangement is independent of any accidental change in the pole strength of the magnet. That piece of apparatus consisting of a circular coil with a magnet suspended at its centre and having a divided circle so that the deflections of the needle can be obtained, is called a **tangent galvanometer.**

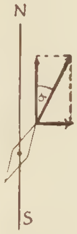


FIG. 121.

In other types of instrument the deflection is observed either by means of a light pointer which moves with the needle, or else by means of a beam of light reflected from a mirror attached to the magnet (mirror galvanometer). When the needle is deflected, the reflected beam moves across a scale. Sometimes the coil is suspended freely in the field of a powerful magnet. When a current is passed through the coil it will rotate. Such instruments are known as **suspended coil galvanometers.**

Finally there are instruments known as **string galvanometers**, which can be used to measure a current. A very thin wire forms a conductor and this is stretched between the poles of a strong magnet. If a current is passed through the wire it will be acted upon by forces the direction of which is opposite to that given by Ampère's rule for a magnet. The small displacements of a very fine wire, sometimes as thin as 0·001 mm. in diameter, can be measured by means of a microscope.

Relative and Absolute Measurements of Current. If it is only necessary to compare different currents with one another, *i.e.* when only the relative values of currents are required, it is sufficient to pass the currents through the same galvanometer. Then, from equation (129), it follows that the ratio of the tangents of the angles of deflection is the same as the ratio of the currents. But if it is necessary to know how many units of electricity have passed through the conductor in one second, we must know the constant C which occurs in that equation. For this purpose it is only necessary to pass a known current on some one occasion through the galvanometer, and observe the corresponding angle of deflection; then all the terms in equation (129) are known except C, which can therefore be calculated. Such a determination has been made. A large condenser of known capacity C is charged many times per second by means of the rapidly rotating and insulated cylinder A, Fig. 122, which has a contact piece B. The condenser is raised to a potential V, and when the contact piece is at b it is discharged through the galvanometer G. In this way the number of units of electricity passing per second through the galvanometer is known, viz. cVn, where c is the capacity, V the potential and n the number of rotations per second. These measurements were first made by **R. Kohlrausch** and **W. Weber** at Göttingen in 1856. The value they found for C was $1/(3 \times 10^{10})$, which is exactly the inverse value of the velocity of light in empty space. This determination has been of the utmost significance in the further development of science, for here, for the first time, a relationship between light and electricity was revealed.

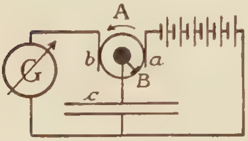


Fig. 122.

From this result **Maxwell** was personally convinced that there are electric waves which travel with the velocity of light, and further, that light waves are really such electromagnetic waves. He then developed the electromagnetic theory of light and to-day we have wireless waves encircling the earth day and night.

Mutual Action between Two Currents. Since a current in a conductor is equivalent to a magnet as far as the excitation of a magnetic field is concerned, it follows at once from what has been said that currents in two conductors will exert a mutual effect upon each other ; in fact the N-seeking pole of one will tend to approach the S-seeking pole of the other. This happens when similar poles point in the same direction or when the currents in the two conductors have the same direction. Hence we have the rule that currents in the same direction attract one another and two conductors tend to move so that the currents flow in parallel directions.

2. The Systems of Electrical Units

The Electromagnetic System of Units. The constant C in equation (123) and those which follow is extraordinarily small. On former occasions we have been able to make such constants disappear from our equations by making them equal to ‘ unity ’. Now, cannot a similar procedure be adopted here ? Strictly speaking, no, because we are no longer free to select the unit since all the variables which occur in the equation have already been completely defined in terms of the units of electric charge and magnetic intensity.

But since the constant has a value which is inconveniently small, so that it is always necessary to work with currents which, in order for the deflections to be measurable, shall be represented by large numbers, it has accordingly been decided to abandon the unit of electricity as already defined ; or rather to use a new unit which can be used whenever it is found that the former unit is inconvenient. This new unit of current strength, I, is so chosen that the above constant C is actually unity. This new unit is called the electromagnetic unit (*E.M.U.*) in contrast to the earlier one which is known as the electrostatic unit (*E.S.U*).

In order to define this new unit we return to equation (125). Let us consider a circular circuit whose radius is 1 cm. and calculate the effect due to a portion of the coil 1 cm. long. According to the equation mentioned above, the force on a unit magnetic pole at 1 cm. distance from the wire is CI. The particular current which must be passed so that this force is equal to 1 dyne is the new unit current.

A current is said to have unit strength (in the *E.M.S.*) when it is such that if it flows in a circuit of unit radius, it exerts a force of 2π dynes on a unit magnetic pole placed at the centre of the circle.

Experiment has shown that this unit of current is 3×10^{10} times larger than the unit of current based on the electrostatic system of units.

The new unit has dimensions which are different from those of the former unit. Let us insert the dimensions of each term in equation (123). Since

$$f = [mlt^{-2}],\ ds = [l],\ M = [m^{\frac{1}{2}}l^{\frac{3}{2}}t^{-1}],$$

it follows that the dimensions of the *E.M.U.* of current are

$$I_M = [m^{\frac{1}{2}}l^{\frac{1}{2}}t^{-1}].$$

Thus, for a current of electricity, the ratio of the electrostatic to the electromagnetic *dimensions* is

$$\frac{[m^{\frac{1}{2}}l^{\frac{3}{2}}t^{-2}]}{[m^{\frac{1}{2}}l^{\frac{1}{2}}t^{-1}]} = [lt^{-1}] = [v].$$

But if σ and ω are mere numbers denoting the *same* current on the electrostatic and electromagnetic systems respectively, then

$$\sigma[m^{\frac{1}{2}}l^{\frac{3}{2}}t^{-2}] = \omega[m^{\frac{1}{2}}l^{\frac{1}{2}}t^{-1}]$$

$$\therefore \frac{\sigma}{\omega} = \left[\frac{1}{lt^{-1}}\right] = \left[\frac{1}{v}\right]$$

It should be observed that since σ and ω are the magnitudes of the same quantity of electricity, their ratio is the *inverse ratio of the size of the units, i.e.*

$$\frac{\text{The } E.M. \text{ unit of current}}{\text{The } E.S. \text{ unit of current}} = v.$$

The numerical value of v is 3×10^{10} cm. per second, *i.e.* it is equal to the velocity of light in vacuo.

Having defined the unit of current in electromagnetic units, the remaining electrical units must also be expressed in terms of the same units ; in fact there are two complete systems of electrical units. Let us mention the following as an example. In order for the heating effect due to a current to be still measured by the formula $W = I^2R$ and the mechanical unit of work to remain the erg, then when I is to be measured in terms of a unit 3×10^{10} times larger so that the numerical value of a definite amount will be 3×10^{10} times smaller in this unit, we shall have that the numerical value of the resistance must be 9×10^{20} times larger, or the unit of resistance in the electromagnetic system is 9×10^{20} or C^2 times smaller than in the electrostatic system.

By using Ohm's law it follows that the *E.M.U.* of potential is 3×10^{10} times smaller than the *E.S.U.*

This new stipulation implies that the magnetic effect due to a current of one *E.S.U.* is very small.

The Practical System of Units. Before the electrical units had been defined according to the above two systems of units, yet a third system had been instituted. It is found to be more convenient in actual practice, so that to-day we have a third system called the practical system of units. The practical unit of current is **one-tenth of the *E.M.U.* of current** and is called the **ampère.** The practical unit of potential is equal to 10^8 *E.M.U.*, or $\frac{1}{300}$ *E.S.U.*, and is called the **volt.**

Again, by applying Ohm's law, it follows that the practical unit of resistance is equal to 10^9 *E.M.U.* of resistance, or $\frac{1}{9 \times 10^{11}}$*E.S.U.* The practical unit of resistance is called the **ohm.** The ohm is equal to the resistance of a column of mercury 1 mm.^{2} in section and 106·3 cm. long.

3. ELECTROMAGNETIC INDUCTION

Electromagnetism. If it is true that there is a reciprocal action between magnets and currents in conductors, then the small molecular magnets in a piece of iron must be rotated by the electrical forces of some external field so that the currents tend to be parallel to one another ; *i.e.* all the poles point in the same direction. In this way the strength of the magnetic field must be increased considerably.

When a piece of iron is surrounded by a coil of wire carrying a heavy current, a very powerful magnet is produced. These are called **electromagnets** and they play a considerable rôle in the electrical industry. From the delicate diaphragm of a telephone to large motors, their motion is produced by means of an electric current and a magnet. Between the poles of an electromagnet field strengths up to 50,000 lines per cm.^{2} can be attained.

Induction due to Currents. Whilst investigating the effect of currents on circuits in which there was no current **Faraday** made the following discovery in 1831. He observed that whenever the current was started or broken in the first (primary) circuit, the needle of a galvanometer in the second (secondary) circuit was deflected. Fifty years later it was not only possible to move suspended magnetic needles, but powerful machinery could be set in motion by this effect.

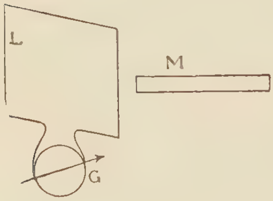


FIG. 123.

The laws of **induction,** as Faraday termed this phenomena, are of the greatest import in electrical engineering, but they have not been of much service in revealing the structure of matter. Accordingly, we shall only discuss them briefly.

Let L be a closed circuit in which there is included a galvanometer G, Fig. 123, to serve as a detector for a current which may be excited in L. The galvanometer will show a deflection every time a magnet M is made to approach the circuit, so that the number of lines of magnetic force threading the coil is changing. When this number is increasing the current passes in a direction opposite to that which flows when the number is decreasing. A N-seeking pole causes the current to pass in a direction opposite to that created by a moving S-seeking pole. The current only lasts whilst the number of lines of force is changing. Now all processes whereby a current can be induced are equivalent, when the same change in the number of lines occurs. The approach of a north pole has the same effect as a receding south pole, or when a circuit is made to approach a stationary north pole. Induced currents can be created by rotating a magnet in the vicinity of a circuit, if the number of lines of force is thereby changed. In place of a magnet a conductor carrying a current can be used. Induced currents will be produced whenever this circuit approaches another, or is rotated, or even if the current is increased or diminished ; finally, they will also be produced if the coils comprising the circuit in which the current is induced (*i.e.* the secondary) are pushed together, or pulled apart.

The direction of the induced current is always such that the motion is checked by the mutual action between the magnetic field and the induced current. In consequence of this the motion is always associated with the expenditure of work and this work must supply the energy for the induced current. By the approach of a N-seeking pole the induced current is such that the near side of its equivalent magnetic shell is a N-seeking pole, but by the removal of a N-seeking pole a S-seeking pole is created, the effect of which tends to prevent the motion.

From these facts, however, it must not be concluded that the laws of magnetic induction can be deduced as a necessary consequence of the laws of energy. That is certainly not so, for even if nothing happened in a circuit whenever a magnet was caused to move in its vicinity, the principle of the conservation of energy would still be true. But when it has been established experimentally that a current is induced, then one can deduce from the above principle the fact that the current will have a definite direction.

The laws concerning the magnitude of an induced current are very simple. Initially the circuit becomes the seat of an electromotive force which causes the electricity to move according to Ohm's law. The strength of the current is equal to the ratio of the induced *E.M.F.* to the total resistance of the circuit. We therefore need only concern ourselves with the magnitude of the induced *E.M.F.* This is obtained as follows. Let us assume that the total change in the number of lines of force threading the circuit is ΔN, so that ΔN represents the difference between the number of lines threading the circuit in its initial and final states. Let the change take place in time Δt. For the sake of simplicity we shall assume, at first, that the change takes place regularly during this time, so that the change per unit time is constant and equal to $\frac{\Delta N}{\Delta t}$. **The induced electromotive force is given by**

$$V = -\frac{\Delta N}{\Delta t} \quad . \quad . \quad . \quad . \quad . \quad (133)$$

This induced *E.M.F.* is only maintained whilst the change in the number of lines is proceeding. If, for example, AB and CD (Fig. 124), are two railway lines, then these will be threaded by the lines of force due to the earth's magnetic field ; over an area which is not too great one may assume that the number per cm.^{2} is constant. If A and C are connected to a galvanometer G whilst a carriage is moving along the rails with uniform velocity, then the axle of the carriage will complete the electric circuit. The motion of the carriage will cause the number of lines of force threading this circuit to change regularly and continuously. If Z is the number of lines of force due to the vertical component of the earth's magnetism which thread a horizontal area of 1 cm.^{2}, whilst b is the distance apart between

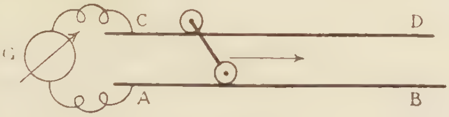


Fig. 124.

the two rails, and v is the velocity of the carriage, then since the area increases by an amount bv every second, the increase in the number of lines of force associated with the circuit is $Z.b.v.$ Hence, as long as the carriage moves uniformly the *magnitude* of the induced *E.M.F.* will be

$$\frac{\Delta N}{\Delta t} = V = Zbv.$$

If R is the resistance of the circuit, then the strength of the current in it is

$$I = \frac{V}{R} = \frac{\Delta N}{\Delta t} \cdot \frac{1}{R} = \frac{Zbv}{R}. \quad . \quad . \quad . \quad (134)$$

We have assumed that $\frac{\Delta N}{\Delta t}$ is constant; the above equation shows that the current will also be constant. The current will be maintained for as long as the *E.M.F.* persists. The quantity of electricity which flows in time Δt is $Q = I\Delta t$; inserting the value of I from equation (134) we have

$$Q = I\Delta t = \frac{\Delta N}{\Delta t} \cdot \frac{1}{R} \cdot \Delta t = \frac{\Delta N}{R} \quad . \quad . \quad . \quad (135)$$

This equation is not a function of the time. The total quantity of electricity which is produced by the inducing agent is independent of the time over which the charge occurs, the velocity of the charge, and only depends upon the total charge in the number of lines threading the circuit and the resistance.

If, as generally happens, the change of flux is not regular, we imagine that the whole change is divided into a large number of small subsidiary changes, which we assume to be constant during the short interval of time in which they occur. Let these small changes be $n_1, n_2, n_3 \ldots$, so that $\Delta N = n_1 + n_2 + n_3 + \ldots$; let these changes take place in times $t_1, t_2, t_3 \ldots$, so that $\Delta t = t_1 + t_2 + t_3 + \ldots$ If the quantities of electricity associated with each single change are $q_1, q_2, q_3 \ldots$ we have, according to equation (135),

$$\begin{aligned} q_1 &= n_1/R \\ q_2 &= n_2/R \\ q_3 &= n_3/R \\ &\cdot \quad \cdot \quad \cdot \quad \cdot \\ &\cdot \quad \cdot \quad \cdot \quad \cdot \end{aligned}$$

Adding all these equations together, we obtain for the total quantity of electricity which passes, viz.,

$$Q = \Sigma q = (n_1 + n_2 + n_3 + \ldots)/R = \frac{\Delta N}{R}.$$

Hence, in general, when the charge in flux occurs irregularly we obtain the same result as before, viz. that the total quantity of electricity is equal to the total change in flux divided by the resistance of the circuit.

The *E.M.F.*'s which are induced in the times $t_1, t_2, t_3 \ldots$ are $\frac{n_1}{t_1}, \frac{n_2}{t_2}, \frac{n_3}{t_3} \ldots$ respectively, so that, when t is small, and the change of flux is irregular, it is preferable to write

$$V = -\frac{dN}{dt} \quad . \quad . \quad . \quad . \quad . \quad (136)$$

In this equation V is to be expressed in electromagnetic units.

Frequently the conductors are of such a shape that the magnetic lines of force cut the circuit several times in the same direction, *e.g.* this happens when a current is produced in a solenoid. Then the change in flux is to be reckoned every time the lines thread each winding of the coil. If n is the number of windings in the solenoid, the magnitude of the induced *E.M.F.* is given by

$$V = -n\frac{dN}{dt}.$$

Self-induction. Another important deduction can be made from the laws of electromagnetic induction. The lines of magnetic force which are produced by the current changing in any one winding of a solenoid, will thread each of the other windings. If therefore a current is created in the circuit by introducing a battery into it, an electromotive force is induced in the same circuit which *opposes* the *E.M.F.* due to the battery itself. The consequence of this is that the current does not assume its final value at once. The strength of this counter-effect will depend essentially upon how often each line of force cuts the coil and whether all the lines of force cut all the other windings. The first condition depends upon the number of windings in the coil, whilst the second depends upon their distance apart. If all these effects are denoted by a symbol L, then the *E.M.F.* which is induced in the circuit when the current is changing at the rate $\frac{\Delta I}{\Delta t}$ is given by

$$V = -L\frac{dI}{dt} \qquad . \quad . \quad . \quad . \quad (137)$$

L is called the **coefficient of self-induction,** or the **self-inductance** of the coil. It only depends upon the number and position of the various parts of the winding with respect to one another. The self-inductance prevents a current from reaching its final value at once, when the current is made in a circuit. When the circuit is broken the self-inductance tends to prevent the current from becoming zero immediately. It is the cause of the small arcs which are seen at the switch when a circuit is broken. The distance across which the arc due to the induced *E.M.F.* passes, is, according to equation (137), greater when the current which is broken is greater; it is also greater when the current is broken more rapidly, *i.e.* the smaller Δt becomes, and finally, it is increased when the number of lines of force is greater, *i.e.* when L is large. This last number, as we have already seen, can be made very big by introducing an iron core into the solenoid. When a heavy current is passed through the coils of such a large electromagnet as can be built to-day, several minutes may elapse before the current assumes its maximum value.

Electrical Oscillations. Briefly, it may be said that self-induction tends to prevent any change from taking place in a current. The effect which self-induction has on electricity in motion is very similar to the part played by inertia when ordinary masses are set in motion. Both tend to prevent any change in the state of motion taking place. For example, let A and B, Fig. 125, be two condensers, one plate of each between joined by means of a coil to spark-gap F. Let the condenser A be charged so

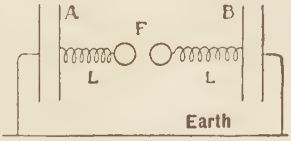


Fig. 125.

that a spark can pass across this gap F. While this happens the condenser A begins to discharge itself through the coil L, the electricity passing to the second condenser B, which we assume to be uncharged initially. The strength of the current in the first place is not given by Ohm's law, $I = V/R$, where V is the potential of the condenser A, for the electric current produces an opposing $E.M.F.$ which is equal to $L \cdot \frac{dI}{dt}$, so that the effective potential difference is

$$V - L\frac{dI}{dt} = IR.$$

After a short time the condenser A is discharged. In consequence of the self-inductance of the circuit, however, the current does not cease at this instant but continues to flow in the same direction, so that B acquires a charge at higher potential. This increase in the potential of B will soon bring the current to zero value when B is charged just like A was initially. The only difference is that the magnitude of the charge is less, since during the time that the current persisted some of the electrical energy will have been converted into heat in overcoming the resistance due to L and F. The discharge will now take place in the reverse direction, after which these to-and-fro discharges will occur continually until all the electrical energy has been converted into heat.

The similarity between this process and the motion of a pendulum is at once apparent, for the self-induction plays the same part as inertia does when a body is moving. When the potential difference has ceased to exist the self-induction causes the motion not to be interrupted, but, shooting beyond the equilibrium position, a new charge and potential are produced, first on one side and then on the other.

On certain occasions it can even happen that the self-induction and inertia affect the motion of the same body. If we were not acquainted with the idea of inertia, we might not have been able to distinguish it from self-induction. We might not have been able to say whether a certain phenomenon depended upon the self-induction or upon the inertial mass of the body. And if it were then necessary for us to make some inference regarding the mass of the body we should not be able to say with certainty whether we were dealing with an inertial mass, or a moving electric charge. We shall soon find ourselves actually confronted with this difficulty.

An Explanation of Electromagnetic Induction. If we now pause for a moment and meditate upon the things which have been said concerning electromagnetic induction, we shall perhaps soon perceive that all our statements have been merely descriptive. In order to have an explanation of the phenomena it would be necessary that one should deduce the motion of the electricity in a circuit produced by a change in magnetic flux, from the properties of electrons and molecular magnets.

We are indebted to **H. A. Lorentz**, a Dutch physicist whose researches concerning the electron theory place him in the front rank of scientists, for an experiment in this direction. As a starting point he assumed the fact that a conductor carrying a current is deflected when placed in a magnetic field with which he coupled the following information due to **Riecke**; it will be remembered that he had shown that the motion of electricity in a conductor was due to the negative charged electrons, whereas the corresponding positive charges were bound to the metallic atoms and took no part in the process. Lorentz then showed that if such a conductor

L [*cf.* Fig. 123] is moved through a magnetic field the motion of the positive as well as the negative charges constitute a current so that both will experience an action upon them due to the magnetic field according to equation (123). But whereas the positive charges are fixed to metallic atoms and cannot move, the electrons are free to move. They therefore follow the force due to the magnetic field, and it is this motion of electrons which forms the induced current. The more exact mathematical theory which Lorentz developed showed that the direction of the electrons agrees with the observed direction of the induced current. The above theory is an explanation of phenomena due to self-inductance, in so far as it explains it in terms of other phenomena concerning the mutual effects between a current and a magnet.

CHAPTER V

ELECTROLYSIS

1. Fundamental Facts

ON several previous occasions we have accounted for certain laws in terms of various hypotheses, without giving any indication of the facts which have led us to form such theories. We refer especially to the application of the electron theory in explaining certain phenomena in electricity. The reader must remember, however, that we have considered only the laws as being established, and that the justification of the explanation given in terms of electrons has now to be shown. The connexion between matter and electricity, to which we have already made reference, appears very clearly for the first time in the phenomenon known as **electrolysis.** We shall first proceed to enumerate the facts briefly, and afterwards state the conclusions which have been drawn from them.

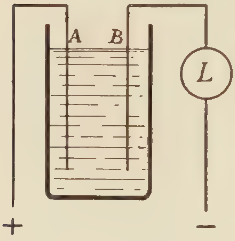


Fig. 126.

Fundamental Phenomena. The following is a well known class-room experiment. A beaker is filled with distilled water into which dip two platinum wires A and B; these two wires do not touch one another (Fig. 126). This apparatus is connected in series with a lamp L and to the mains supplying either direct or alternating current. The lamp does not glow; the resistance of the water is so great that no current passes through it. Pure water is not a conductor for electricity. If a glass rod is dipped in strong sulphuric acid so that it is thoroughly wetted, and then rinsed in the water, the lamp instantly glows brightly. Water containing a trace of acid is a good conductor. Further researches have shown that water can be made conducting by the addition of any acids, bases or salts. In order to obtain comparable results due attention must be paid to the following points: the quantity of water used must always be the same, the distance apart of the two wires must be constant, and the potential across the cell must also be constant. In addition to these, the same mass of the substance under investigation must be introduced into the water, or, as may be expected from former considerations to be better, the quantity of substance added should always contain the same number of molecules, *i.e.* the quantity should be proportional to the molecular weight of the dissolved substance.

If the conductivity of these solutions is investigated no very surprising results are obtained at first; the conductivity is a characteristic for each substance, but varies between wide limits for different solutions. In general it is found that those acids which are termed strong are the best conductors, whereas weak acids only conduct the current feebly. When other substances, which do not belong to any of the above three classes, are dissolved, *e.g.* sugar, the water does not become a conductor. From these facts we have to conclude that it is the dissolved substance which is mainly, if not entirely, responsible for rendering the water conducting. If such a view is correct, we may further suppose that the conductivity of a solution will be greater when the quantity of acid or base dissolved in it is increased. This may be tested very easily as follows. A dilute solution of sulphuric acid is allowed to fall a drop at a time into pure water; the increase in current which accompanies the addition of each drop will take place by equal amounts, if all other factors are kept constant. The experiment may also be performed using a dilute aqueous solution of a salt. In all instances this proportional increase only occurs providing that the quantity of dissolved substance has not exceeded a certain amount.

Whilst performing such experiments as the above there is another phenomenon which soon appears. We refer to some happenings which occur near to the two platinum wires which carry the current to the solution. If, for example, hydrochloric acid has been added to the water, bubbles of gas appear to rise from the two wires whilst the current is passing. When the solution is contained in a vessel shaped like a U-tube so that the two electrodes are situated one in each limb, as in Fig. 127, then the well-known smell of chlorine gas is soon noticed near to the electrode connected to the positive pole of the battery. The gas which appears at the other electrode is odourless, but if it is collected and burnt it is found to be hydrogen. Chlorine and hydrogen are the two elements which form hydrogen chloride. Hence we see that the hydrogen chloride in the water not only conducts the current, but that during this process it is decomposed into its two constituents.

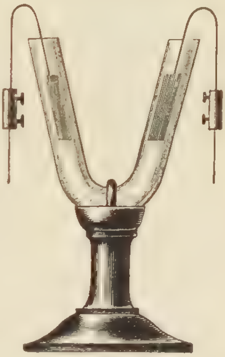


Fig. 127.

Faraday's Laws of Electrolysis. **Faraday** was the first person to investigate these phenomena carefully; his experiments were made during the years 1833–1835. He introduced the notation which is usually adopted. The substance decomposed by the current is the **electrolyte**, the decomposition take place by **electrolysis**, and the two metals which convey the current to the solution are the *electrodes*. The electrode which is connected to the positive pole of the battery is the *anode*, the other is the *cathode*. The substance which appears at the anode is the *anion*, the other is the *cation*.

Faraday also discovered that the decomposition of substances could be obtained without the use of an aqueous solution; it was only necessary to heat the substances so that they were molten. For example, potassium iodide melts at 630° *C*. If it is placed inside a vessel and heated until it melts, it conducts electricity without the addition of water. It is decomposed by the current, iodine appearing at the anode, and potassium at the cathode.

By the quantitative determination of the current used and the amount of substance appearing at the electrodes Faraday was led to discover the two **Fundamental Laws of Electrolysis.** The first one states that **the amount of chemical decomposition is always proportional to the quantity of electricity which has passed through the electrolyte.** This means that when the current is constant the quantity of substance deposited per second is also constant. But if the current is doubled, then, in the same time, the quantity of substance deposited is also doubled.

In the electromagnetic system of units we have hitherto only defined the strength of the current and not the quantity of electricity. Hence we must now determine the quantity of electricity in terms of the current, a procedure exactly opposite to that which we adopted in connexion with the electrostatic system of units. We can do this by stipulating that the unit of electricity shall be that quantity which passes per second along a conductor carrying a current of unit strength. Equation (116) is again valid. The quantity of electricity associated with a current I flowing for t secs. is $Q = It$. Since in the present example the quantity of substance liberated is greater according to the measure in which the current, or its time of duration, is increased, *i.e.* the quantity of substance appearing at the electrode is proportional to It. Faraday could quite correctly have expressed his first law by saying that the quantity of substance liberated is proportional to the quantity of electricity which has passed. The mass of substance deposited can be used as a means of measuring a quantity of electricity; later on we shall perhaps have occasion to resort to such a process.

The second law of Faraday compares the mass of different substances deposited when the same quantity of electricity is used. The experiment is very easily performed. The same current must be passed through several electrolytic cells arranged in series (Fig. 128). By introducing a battery at any point in the circuit the condition that the same current shall pass through each cell is automatically fulfilled. It is therefore only necessary to compare the quantities of substance deposited in the various cells. Faraday arrived at some very important conclusions when the quantity of substance deposited was expressed, not in grams but in gram-atoms. In the first place he showed that for a whole series of elements, the chief of which was hydrogen, the mass liberated contained the same number of atoms. But with other substances the number of atoms liberated was exactly one-half, for others one-third, of that liberated when substances belonging to the first series were employed. He was therefore able to arrange all elements into three classes according as the number of gram-atoms liberated was equal to that of hydrogen, or one-half or one-third that value. On further considering the matter he discovered that they could all be classified according as they were mono-, di-, or trivalent elements. He therefore concluded that a divalent atom gives up a charge of electricity when it is deposited which is double that given up by a monovalent atom; for a trivalent atom the charge will be threefold. If one compares these statements with the fact that a divalent atom may combine with two atoms of hydrogen, a trivalent with three, etc., then these ideas may be expressed as follows, when we do not concern ourselves with the number of atoms in the deposited substance, but rather ask how many atoms of hydrogen will combine with (or be replaced by) them. The answer is as follows: **the mass of material liberated by a constant amount of electricity can also combine with the same number of atoms of hydrogen.** The capacity which an atom possesses in virtue of which it is equivalent to one hydrogen atom was called by Faraday its **valency,** so that the above result may be expressed more simply thus:

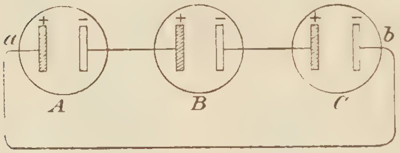


Fig. 128.

The mass of an element liberated by a definite quantity of electricity is proportional to its chemical equivalent

(chemical equivalent = atomic weight ÷ valency).

Voltameters. Proceeding in the converse manner, the quantity of electricity which has passed through an electrolyte can be determined from a knowledge of the mass of material deposited at an electrode; if the current is known to be constant, its strength can then be calculated. Instruments used for this purpose were called by Faraday voltameters, a term which is still retained. Faraday designed the first voltameter (Fig. 129), and by measuring the quantity of hydrogen and oxygen collected in it, he was able to estimate the magnitude of the current. In order to compare one current with another this apparatus alone is sufficient, but if one wishes to express the current in amperes, the voltameter must be calibrated, *i.e.* on some one occasion the strength of the current must be determined accurately by some other means, and the amount of substance liberated at the electrode ascertained under the same conditions. The amount liberated by unit current in one second is then calculated.

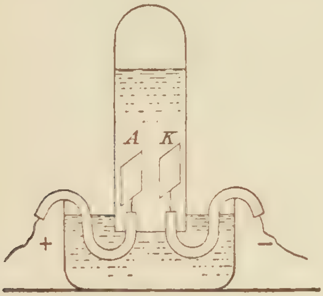


FIG. 129.

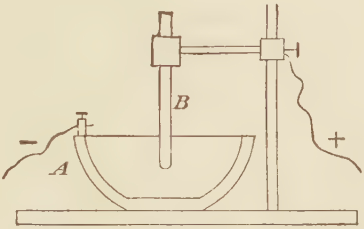


FIG. 130.

Later on it was found that the most reliable measurements could be made if solutions of silver salts were used and the mass of silver deposited at the cathode determined. The current entered a solution of silver nitrate at the anode B, Fig. 130, which consisted of a thick rod of silver, and returned to the battery by way of a wire connected to the platinum dish A which formed the cathode. The silver was deposited on this dish so that its mass could be determined. This mass is now used to define the ampere: **it is that constant current which deposits in one second 0·001118 grm. of silver.** Since a gram-atom of silver amounts to 107·88 grm., an easy calculation shows that in order to deposit a gram-atom of silver, or in fact the gram-equivalent of any element, 96,495 ampere-seconds are necessary. Now an ampere-second is called a coulomb and the number

$$F = 96{,}495 \text{ coulombs} \qquad (138)$$

is known as **Faraday's constant.** In the electrostatic system of units this must be multiplied by 3×10^9, so that

$$F = 28{\cdot}95 \times 10^{13}\ E.S.U.$$

In his commemorative address on Faraday in 1881, **Helmholtz** (1821–1894) expressed these discoveries in such a way that we can form a very clear mental picture of them. Since 1 gram-atom of any monovalent substance conveys the same quantity of electricity to the electrode, and all gram-atoms contain the same number of atoms, it follows that each individual atom of a monovalent element carries with it the same amount of electricity ; a divalent atom will carry twice this charge, and a trivalent three times. The quantity of electricity associated with each atom of a monovalent element is therefore independent of the particular kind of atom. From this behaviour, Helmholtz concluded that even electricity itself is atomic in structure and that there were tiny charges which entered into combination with the elements. A smaller quantity than this could never be found on any element ; if the charge at any time were found to be greater, then its magnitude was not arbitrary, but just two, three, or more integral times that amount. In reality these were exactly the reasons from which Dalton concluded that matter had an atomic structure : in all chemical reactions a certain minimum quantity of reagent is necessary—it can never be less than this minimum amount. In the present instance we have a definite minimum quantity of electricity, and no charge can ever be less than this minimum. We are even able to estimate the magnitude of this minimum charge. According to the above, each gram-atom has a charge amounting to $28{\cdot}95 \times 10^{13}$ *E.S.U.*, and since each gram-atom contains $60{\cdot}5 \times 10^{22}$ atoms, each atom must carry a charge of

$$28{\cdot}95 \times 10^{13}/60{\cdot}5 \times 10^{22} = 4{\cdot}77 \times 10^{-10}\ E.S.U.\ ;$$

this charge is considered to be the atom of electricity. We call this charge e, and write

$$e = 4{\cdot}77 \times 10^{-10}\ E.S.U. \qquad . \quad . \quad . \quad (139)$$

That this small quantity of electricity is carried by each atom of hydrogen to the electrode is a fact established by observation, or at least a fact which is derived by simple division from something which has been observed. **That the reason for this constant occurrence is founded in the atomic structure of electricity was a hypothesis due to Helmholtz.** Later investigations have been so thoroughly in accord with this hypothesis that its validity is no longer doubted. The further reasons which are in favour of this assumption will soon be learned.

In this fact we also find additional evidence to confirm the view expressed earlier on, viz. that electricity is very closely related to matter. In particular, we discern a connexion with that chemical property of a substance which we usually refer to as its valency. We now see that those elements which combine chemically carry opposite charges with them, and it is on account of these charges that they enter into chemical union. The capacity which an element has to combine with a definite number of monovalent atoms is due to the fact that it carries this same number of elementary charges of electricity.

Secondary Processes. Faraday soon discovered that the products of electrolytic decomposition are not always to be observed directly at the electrodes. It often happens that the liberated elements react immediately with the material of the electrode or the electrolyte, and it is the products of these reactions which ultimately appear at the electrodes.

When an aqueous solution of hydrogen chloride is electrolyzed, hydrogen is formed at the cathode and chlorine at the anode. Both these substances are liberated without any further reaction taking place. It must be observed, however, that these gases are partly soluble in water so that all the gases produced are not actually set free. But if the electrolyte is saturated with these gases, the amounts of gas liberated conform with the values deduced theoretically.

The process is somewhat different when a solution of sodium chloride is decomposed by the electric current. At the anode, chlorine is again set free, but the metallic sodium which is formed at the cathode reacts at once with the water, which it decomposes according to the well-known formula,

$$2Na + 2H_2O \rightarrow H_2\uparrow + 2NaOH.$$

The caustic soda remains in solution but the hydrogen appears as a gas at the electrode, and it should be observed that the amount set free, namely the two atoms, corresponds exactly with the two atoms of sodium which should be set free from the solution according to the laws of Faraday. Moreover the composition of the electrolyte will gradually change, for caustic soda will collect in the vicinity of the cathode.

When there is sulphuric acid, or any sulphate, such as sodium sulphate, for example, in the water, then the events which take place at the cathode are exactly the same as in the case of hydrogen chloride solutions, *i.e.* hydrogen is liberated directly from the sulphuric acid, or, as with sodium sulphate, from the liberated sodium which reacts with the water giving rise to the formation of the same amount of this gas.

At the anode, however, the SO_4 group of atoms which are set free reacts with the water, liberating two atoms of oxygen according to the equation,

$$2SO_4 + 2H_2O \rightarrow 2H_2SO_4 + O_2\uparrow.$$

The quantity of oxygen liberated is again in accord with Faraday's laws. When sulphuric acid is subjected to electrolysis this latter substance (H_2SO_4) is always being formed at the anode, so that the quantity of acid in solution remains unchanged. But at the two electrodes, hydrogen and oxygen are set free continuously, *i.e.* the corresponding amount of water is always being decomposed. The whole process is often referred to as the decomposition of water by means of an electric current. It is to be observed, however, that the sulphuric acid is only being formed at the anode. If the vessel containing the electrolyte is in the form of a long tube so that the tendency for its composition to be constant is hindered (a porous membrane can be used to prevent the liquid from having a uniform composition), then the process at the two electrodes can be investigated separately; the results obtained confirm the views which have already been expressed. If sodium sulphate, Na_2SO_4, is the dissolved substance, free acid is formed at the anode, caustic soda at the cathode, but the liquid in the middle remains unchanged.

When a salt whose metallic radicle does not react chemically with the electrode, *e.g.* copper sulphate, $CuSO_4$ is decomposed, then the copper is merely deposited on the cathode. Sometimes the metal forms a coherent deposit on the metallic cathode; the process is then used in industry for plating various articles with copper, gold, nickel, etc. For this purpose platinum, which is not attacked by the products of the electrolysis, is not used as the anode, but the same metal as is to be deposited, viz. copper, nickel, or gold. Then, for example, the SO_4 radicle, which is liberated, reacts with a copper atom from the anode combining with it to form $CuSO_4$ again. The metal from the anode is therefore continually dissolved, but the amount of salt in solution is invariable. It is as if the metal had just passed from anode to cathode. But if the two electrodes are separated from one another by a considerable distance, then it is quite easy to show that the amount of copper sulphate at the anode is continuously increasing, whilst that at the cathode diminishes.

Hittorf's Transport Numbers. The variations in concentration which occur in the neighbourhood of the electrodes were investigated during the years 1853–1859 by **Hittorf** (1824–1914); these researches enabled him to obtain an insight into the processes which were at work during electrolysis. Hittorf prevented the liquids around the electrodes from mixing by inserting pieces of parchment paper in the solution. The whole liquid was thereby divided into three portions, a central one, and two outer ones into which the electrodes were placed.

For the purposes of illustration Fig. 131 will serve our needs. The current is passed through three U-tubes arranged in series. After an experiment the tap below the central tube is opened and the liquid removed. This also separates the two liquids around the electrodes from one another. They can also be removed and their contents analysed.

One of the substances investigated by Hittorf was copper sulphate, $CuSO_4$. At first he determined the quantity of copper sulphate in the solution, so that he knew the amount of copper per cm.³. After a current had been passed for some time, the mass of copper deposited on the cathode was determined by means of a balance. The central portion of the liquid was then removed so that the liquids around the electrodes were isolated from one another. Since the central portion of the solution had remained unchanged, it was certain that all the chemical changes, which had been produced by the current, had occurred in those portions of the liquid which were now isolated. Let us first consider the liquid around the cathode. Its volume having been measured, it was then evaporated to dryness so that the mass of solid in it was known. From this information the amount of copper in the solution was easily calculated—let it be b grm. Before the current had been passed suppose that this quantity was c, so that the difference $c - b = d$ says how much copper had been removed from or added to the solution during the time that the current lasted. (In this instance there was found to be a decrease at the cathode.) Moreover the mass of copper a deposited on the cathode is also known; according to Faraday's law this mass is proportional to the strength of the current. If the only thing that occurred during electrolysis was that this mass a was liberated at the cathode, then a would be equal to d, the diminution in copper content of the solution. But this is not so. The diminution in the amount of copper in solution is smaller. Since it is quite certain that a grams have been deposited at the electrode, it follows that the difference

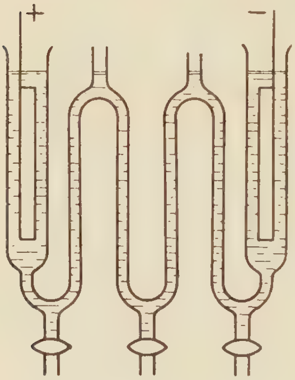


Fig. 131.

$$a - d = a + b - c$$

must be supplied from elsewhere, *i.e.* it must have been carried by the current to some other part of the electrolyte.

In view of these observations, Hittorf made the following suggestions. During the process of electrolysis, a certain quantity of electricity passes from the anode to the cathode. Since the passage of the electricity is closely associated with the liberation of matter at the electrodes, it must be assumed that, in electrolytes, the motion of the electricity is accompanied by a simultaneous movement of material particles. The motion will certainly be caused by the electrical charges which they carry; for these will be attracted or repelled by the charges on the anode and cathode. The current may be due to the fact that the positively charged copper ions move from the anode to the cathode ; or it might equally well be produced by the negatively charged sulphate ions, which travel from the cathode to the anode. For, in this way, the positive charge at the anode would be diminished (it would be spent in neutralizing the negative charge on the SO_4 ions), whilst at the cathode (owing to the positive Cu ions) it would be increased. **Both processes are, however, possible,** for the current may be due in part to the fact that positive Cu ions move in the direction of the current, the remainder being due to the negative SO_4 ions which move in the opposite direction. This problem was solved by Hittorf, who divided the electrolyte into three portions. In whatever way the condition takes place along the central portion of the electrolyte, the amount of substance moving through it towards the anode must be equal to that which moves towards the cathode. This, however, would not be so in the two columns of liquid containing the electrodes ; let us consider the cathode compartment. It certainly lost an amount of copper equal to a. If on the whole, therefore, it had lost less than a grams of copper, the difference must have passed from the other side. If, for example, it had lost nothing, this would indicate that a grams had been transported from the other side and the whole of the current would have been carried by the copper particles. On the other hand, however, if the amount lost had been a, then nothing would have passed over.

But if the loss in copper content for the liquid around the cathode were d grams, then the quantity of electricity associated with $(a - d)$ grm. of copper will have passed through any one section of the tube. But since the actual amount of electricity passing is represented by a, it follows that the fraction of the current carried by the copper ions is $(a - d)/a$.

By determining the copper content of the liquid around the anode, this result can be examined. During the whole process the anode loses a grams of copper which pass into solution. If the increase in copper in the anode compartment does not correspond with this amount, then some must have passed through the partition. Thus, if the increase is d' instead of a $(a > d')$ the difference $(a - d')$ will have moved through the dividing membrane. On the whole a quantity of electricity associated with a grm. of copper will have passed through the membrane ; of this, a fraction represented by $\left(\frac{a - d'}{a}\right)$ will have been carried by the copper ions. If the above hypothesis is correct, then the results of the analysis should show that $d = d'$, for the two above fractions must be equal.

Hitherto we have only discussed the cation, but the same considerations apply equally well to the changes in the concentration of the anion. It is found that another fraction of the current is carried by the anions. The two portions together must give the total current a, or if they are expressed as a fraction of the whole, then the sum of the two fractions must equal unity. Thus, if n is the fraction for the anion, that for the cation is $1 - n$.

Hittorf called these fractions **transport numbers**. They indicate the portions of the current which are carried by the anion and cation respectively.

In general the transport number is greater when the mass of the particle is less. For example, when dilute sulphuric acid forms the electrolyte, the H^+ is the cation * ; its transport number is 0·8, whereas for copper sulphate with Cu^{++} cations the number is about 0·38. The values of these numbers vary somewhat with the concentration and temperature of the electrolyte.

2. THEORIES OF ELECTROLYSIS

Whilst the facts concerning electrolysis were being established, an attempt was also being made to explain the process. Even the discovery of the facts which have been described in the earlier part of this chapter, viz. those which were revealed by Hittorf, was largely due to the conjectures which had been formed gradually with regard to the process of electrolysis. At the time of Hittorf those hypotheses were being developed, which have been confirmed again and again. These hypotheses also indicated the directions in which further researches might be made. A short account will therefore be given of these early hypotheses.

The oldest of these theories, and one which was proposed by **Grothuss** (1805), assumed that the constituents of the electrolyte were torn asunder by the electric forces due to the potential difference between the electrodes. For example, if $NaCl$ was dissolved in water, the negatively charged atom of chlorine was drawn towards the anode, the positively charged atom of sodium, which had been in combination with the chlorine atom, being repelled so that the two atoms were separated. The chlorine gave up its charge to the anode, whilst the sodium atom which was left combined with a neighbouring chlorine atom. This process was continued throughout the electrolyte right up to the cathode, where the remaining sodium atom was attracted to it and then discharged. Half a century later no further light had been thrown on the subject of the decomposition of an electrolyte by an electric current.

In 1857 **Clausius** called attention to the following point : according to the above theory a current will only pass through an electrolyte when the electric force due to the potential difference between the two electrodes has become greater than the attraction between the two parts of the dissolved substance, *i.e.* a certain minimum value must be exceeded before the current will pass. It therefore follows from this theory that the conduction of electricity through an electrolyte could not be governed by Ohm's law, according to which the current must also decrease proportionally with the potential, so that it is only zero when the potential is zero. In practice it is found, for example, that when silver electrodes are used in an aqueous solution of $AgNO_3$, a current begins to pass even when the slightest potential difference is applied to the electrodes. The minimum value, if one wished to speak of such a value, is therefore zero, *i.e.* the forces which hold the particles together are zero ; **in fact the particles are no longer held together at all.** It therefore appears that the two sets of atoms which finally reach the electrodes are already present in the solution from the very beginning, and the current has no task to perform other than that of driving the atoms to the electrodes where they are discharged. It is quite certain that Clausius believed that the bond between the atoms in the solution was broken in consequence of their rapid movements and that they moved in virtue of the collisions which they made with other atoms ; this movement enabled them to combine with other atoms so that any atom was rapidly alternating between a state of freedom and one in which it was in combination with another atom. In order to explain the passage of electricity, it was quite sufficient if a number of the molecules were thus divided, for it was only those atoms having opposite charges, and which Faraday had termed **ions**, that could carry the electric current.

---

* H^+ denotes a *H*-ion with a single positive charge. $SO_4^{=}$ is a sulphate-ion with its two negative charges. Similarly with other ions.

Helmholtz extended these ideas in 1880. He assumed that the ions could remain detached permanently and carry a charge with them. To the objections raised by the chemists, that it is impossible for a sodium atom to be in contact with water, which is decomposed at once by sodium, **Hittorf** had already replied, when he stated that there was a great difference between a sodium atom and a sodium ion. This last-mentioned substance is an atom carrying a positive charge, and by combining with such a charge its properties may be very materially altered, just as for example when it combines with a chlorine atom and forms sodium chloride. Similar remarks may be made with reference to a chlorine ion and a chlorine atom. Helmholtz also showed that the main task of the current was to cause the ions to move through the liquid to the electrodes where they lost their charges.

Electrolytic Dissociation. The above theory was developed more thoroughly by **Arrhenius** (1859–1927) in 1887. He assumed that in an electrolyte **a definite proportion of the dissolved substance is always present as ions** ; this he called the **degree of dissociation** of the electrolyte. It is only these ions which are charged and which are therefore responsible for carrying the current. The degree of dissociation for various electrolytes differs considerably ; for good conductors it is large and may even approach unity, a fact which would indicate that all the dissolved substance is present in the form of ions. For poor conductors the degree of dissociation may be very small. The degree of dissociation is also intimately connected with the density of the solution ; in weak solutions the dissociation is almost complete. The chief value of Arrhenius' theory lay in the fact that his assertions could be used in order to account in a satisfactory manner for other unexplained phenomena.

On page 132 we have referred to the osmotic pressure exerted by all dissolved substances. This pressure is exactly the same as the gas pressure which all the dissolved molecules would exert if they were present as a gas in a volume equal to that of the solution. Like this latter, it is independent of the kind of substance and only depends upon the **number** of molecules in solution, as the general form of the gas equation shows [*cf.* p. 149]. In consequence of this osmotic pressure the freezing point of a solution is lowered, whilst its boiling point is raised ; these effects are likewise independent of the kind of dissolved substance, and only depend upon the number of molecules in unit volume of the solution. A few substances, however, do not behave according to this rule ; these are acids, bases, and salts, *i.e.* electrolytes. With these substances, the lowering of the freezing point, for example, is greater than that calculated from the number of molecules dissolved ; the discrepancy is greatest for good conductors, whereas it is hardly noticeable in the case of poor conductors. Further, with electrolytes like sodium chloride, $NaCl$, for example, which can only dissociate into two ions, the number of particles present can never exceed twice the number of molecules. For other substances, like sulphuric acid, H_2SO_4, which we assume can be decomposed into three ions H^+, H^+ and SO_4^- respectively, the number can approach three times its normal value, but it can never be greater than this. These maximum values are only reached in the case of very dilute solutions.

For solutions which have higher concentrations the results are not so concordant, *i.e.* the degree of dissociation calculated from the electrical conductivity is not the same as that deduced from the depression of the freezing point. In papers published since 1922 **Debye** has accounted for these discrepancies. In particular he has shown for strong electrolytes, *i.e.* those with a large ionic content, that there is a force of attraction between the positive and negative ions, and that a positive particle is more often, and for a longer time, found in close proximity to a negative one than it is to a like particle. He also showed that the electric forces will be extremely big on account of the very small distances between the atoms, so that each will tend to move more slowly ; the consequence of this will be that the conductivity will be diminished, as if the degree of dissociation were less. We therefore speak now of the **effective** degree of dissociation of strong electrolytes. Proceeding in much the same way as Van der Waals did with gases, he has obtained a more accurate equation to represent the conduction of electricity through electrolytes ; the agreement with observation has thereby been made more perfect.

Conduction in Electrolytes. According to the theory of electrolysis as developed by **Arrhenius**, a greater portion of the dissolved substance will be ionized when the solution is made weaker. The electric force arising from the potential difference on the two electrodes will cause both ions to move. Initially this force will impart to them an acceleration but after a very short time a viscous drag will begin to make itself felt. The retarding forces due to viscosity become greater when the velocity of the ion is increased, and this is the reason why the viscous forces soon become equal to the electrical forces. When this has occurred, the particle will no longer be accelerated, but will move with constant velocity through the liquid so long as the electric force remains the same. If, however, this force increases, the viscous forces will no longer be able to hold the particle in equilibrium, so that it will again be accelerated. This increase in velocity will cause the retarding forces on the particle to become larger until the accelerating and retarding forces are again equal ; the particle will then move with constant velocity through the liquid. It is therefore obvious that the ions will move more rapidly as the potential difference between the electrodes is increased ; in other words, Ohm's law will be obeyed.

Kohlrausch's Law concerning Ionic Mobilities. The ideas which had been initiated by **Clausius** led **Kohlrausch** to study the conductivity of electrolytes in 1876. He wished to investigate, in the first place, the part which the individual ions played in electrolysis, and asked himself whether the conductivity due to an ion, for example a sodium ion, was independent of the chlorine or bromine atom in sodium chlorine or bromide with which it had to associate during its motion through an electrolyte. If, according to the view of Clausius which was current at that time, the ions were all independently set in motion by the electrical forces, it was necessary to conclude that the conductivity of a particular ion would be independent of all other ions with which it might be associated. He therefore investigated the conductivities of various solutions placed in the same piece of apparatus, *e.g.* those of $NaCl$, $NaBr$, and NaI. These experiments were followed by others in which the cation was potassium but the anion remained as before. If, now, each ion had its own definite and characteristic conductivity, the difference between the conductivities of $NaCl$ and KCl must be due to a difference between those of Na and K; if these were really characteristic of each ion, then the difference in the conductivities of $NaBr$ and KBr must be just as great as in the above case, since it could only be attributed to the difference between the sodium and potassium ions. Similar remarks may be made with reference to the anions: bromine and chlorine ions should exhibit the same difference whether they were investigated as $NaBr$ and $NaCl$ or as KBr and KCl. These expectations were completely realized so that he was able to express the law that the ions in an electrolyte conduct the current independently of the other ions which are present in the same electrolyte.

Ionic Mobilities. Kohlrausch's law is a direct deduction from observed facts concerning the conductivity of electrolytes. But **Kohlrausch** did not express his law in this form, which is free from any hypothesis. He suggested that the current through an electrolyte owes its origin to the atoms which carry the small charges on them to the electrodes; a positively charged atom travels towards the cathode, etc. Such currents as these are called convection currents. It therefore follows that the conductivity of an ion must be proportional to the velocity with which that ion moves through the electrolyte. The more rapidly the particles move under the action of the impressed electrical forces, the greater is the charge per second which is carried through any cross-section in the circuit. He therefore concluded that the ionic conductivities, which were characteristic of the ions, really depended upon the velocities with which the ions moved through the solution.

This law, however, must be confined to dilute solutions, where it is justifiable to assume that along their path through the electrolyte the ions are only affected by the resistance of the liquid, and that encounters with other ions are so rare that the conductivity is not thereby affected. If such conditions are not realized, then other facts have to be taken into consideration: the work of Debye in this connexion has already been mentioned.

The chief importance of Kohlrausch's law is to be found in connexion with very dilute solutions. He made elaborate experiments with such solutions for many years and in so doing obtained a very valuable insight into the nature of electrolysis; his results are very important in connexion with our main problem, which is an attempt to obtain some knowledge concerning the structure of matter.

Equivalent Conductivities. Kohlrausch then introduced the idea of **equivalent conductivities.** Hitherto, when we have dealt with processes depending upon the number of atoms or molecules present, we have formed the habit of referring our measurements to the same number of atoms or molecules—we introduced the idea of a gram-atom or molecule. We now have to discuss a quantity of matter with which the same amount of electricity is always associated. The gram-atom of a monovalent element always carries a charge F equal to 96,495 coulombs. But a gram-atom of a divalent element would carry twice this amount: similarly trivalent, tetravalent, etc. elements would carry charges which are 3, 4, etc. times as much. In the case of the monovalent OH-ion of water, or the divalent SO_4-ion from sulphuric acid, the associated charges would be F and $2F$ respectively. Hence we form the new idea of a **gram-equivalent,** by which term we understand that mass of ions which always carry a charge of 96,495 coulombs. This mass is equal to the mass of the ions in a gram-ion divided by the valency of the ion. For example, hydrogen is monovalent, so that 1 grm. of hydrogen ions has a charge F coulombs. Oxygen is divalent and 16 grm. of oxygen carry a charge $2F$, so that the amount of oxygen which will only carry a charge F is $16/2 = 8$ grm. In general, if G is the weight of a gram-ion and n its valency, then the electrolytic equivalent, A, of this ion is given by $A = G/n$.

It is now quite easy to understand what the term equivalent conductivity implies. Let us suppose that κ is the specific conductivity of an electrolyte containing c equivalents per cm.^{3}; then the equivalent conductivity is κ/c. (It is obvious that c can assume fractional values, in fact fractions which are very small.) We denote it by Λ, so that $\Lambda = \kappa/c$. If this electrolyte is further diluted and its specific conductivity κ' determined, its equivalent conductivity is now equal to $\frac{\kappa'}{c'}$. The equivalent conductivity is now greater since there will be a greater number of ions in the solution. If the dilution is continued, the value of the equivalent conductivity increases until all the molecules have become dissociated into ions; it has then reached its maximum value, which is not altered if the dilution be further increased. If we use Λ_∞ to denote the value of the equivalent conductivity at infinite dilution, then on both occasions there will be one gram-equivalent of the electrolyte in solution, but it is only in the latter case that all the molecules are completely dissociated into ions which conduct the current; in the first case only a fraction a is dissociated. We therefore have

$$\Lambda : \Lambda_\infty = a \qquad . \quad . \quad . \quad . \quad (140)$$

the **degree of dissociation** for the electrolyte.

3. THE IONS OF ELECTROLYTES

Ionic Mobilities. Let us now assume that we have an electrolyte containing one gram-equivalent per cm.^{3} of the solution. If all is dissociated, then the charge on the ions of any one sign is $F = 96{,}495$ coulombs, but if only a fraction a is dissociated, then this charge is aF.

We shall further assume that the dilution of the electrolyte has been carried to such an extent that the anions and cations do not disturb one another as they move in opposite directions. The law of the independence of conductivities can be applied to this problem. Conductivity is therefore an additive property, that is to say, the total conductivity is equal to the sum of the conductivities of the anions and cations, *i.e.* $\Lambda = \Lambda_a + \Lambda_c$.

We will suppose that a potential difference of one volt is applied to two metal electrodes each of 1 cm.^{2} area and placed at a distance of 1 cm. apart. Let us endeavour to discover the velocities which the ions attain under the influence of the electrical force which is present; let u be the velocity of the anion and v that of the cation. Let us now consider a cross-section in the liquid which is parallel to the electrodes and therefore 1 cm.^{2} in area. The anions convey a charge uaF per second, whilst the cations convey an amount vaF; the total $(u + v)aF$ is the current which can be measured. Since the potential difference is unity, we have, according to Ohm's law, $I = \frac{1}{R} = \kappa$, so that the current is also a measure of the conductivity; in this instance we are dealing with an electrolyte whose equivalent conductivity is Λ, so that $\alpha F(u + v) = \Lambda$. If the dilution is such that all the ions have become dissociated, we have $\alpha = 1$, and therefore $\Lambda_\infty = F(u + v)$; hence, by measuring Λ_∞ we have a knowledge of $u + v$, which is $\Lambda_\infty F$. Moreover, the transport number gives the value of $n/(1-n)$, which is also equal to u/v, so that from the two values $u + v$ and $u : v$ the two velocities with which the ions move in a solution under a potential difference of one volt per cm. can be computed. As an example we may refer to sulphuric acid in which $\Lambda_\infty = 388$; taking $n = 0{\cdot}2$ and $1 - n = 0{\cdot}8$, we have $u : v = 1 : 4$, so that $u = 0{\cdot}0008$ and $v = 0{\cdot}0032$ cm. per sec.

In this way we have actually obtained the true values of the velocities with which the ions move through the solution under the conditions stipulated. The values are very small, but this fact is to be attributed to the smallness of the effective forces and the mechanical resistance of the solution. When the potential difference is applied to the electrodes, the ions are set in motion at once. Those which lie near to the electrodes, let us say the anode, are deposited immediately, whilst those which are present in the more distant parts of the solution move with this velocity through it. Now it is not only the approaching anions which are deposited, but also the particles which belonged to the receding cations. For it is only in this way that the value of the current which goes to the electrode attains the value of the total current which passes through any cross-section of the electrolyte. The above remarks may perhaps help us to realize how far we have progressed in our knowledge concerning electrolysis.

The Mechanical Forces. Finally, Kohlrausch has shown how one may calculate the forces which operate on 1 gram-equivalent under the conditions which have been assumed. When 1 coulomb of electricity moves under a potential difference of one volt the work performed amounts to 1 joule; hence, when 1 gram-equivalent, with its F coulombs, moves under similar conditions, the work done is 96,495 joules, which is approximately equal to 10^{12} ergs $= 10^9$ grm. cm. $= 10^6$ kgrm. cm. Since this work is done whilst the charge moves through 1 cm., it follows that the force is equal to the weight of 10^6 kgrm., or more exactly 984,000 kgrm.

This force imparts to a gram-equivalent of hydrogen a velocity of 0·0032 cm. per sec. In order therefore for this velocity to become unity, the necessary force would be $984{,}000 \div 0{\cdot}0032 = 300$ million kgrm.-wt. For other ions having a still smaller velocity this force would be correspondingly larger. The magnitude of this force is simply enormous.

But if we now calculate the force which is required to impart unit velocity to a single hydrogen ion, we must remember that 1 grm. of hydrogen constitutes its gram-equivalent, which therefore contains $60{\cdot}5 \times 10^{22}$ ions. If we reckon the force in mgrm.-wt., viz. 300×10^{12}, then by dividing by 60×10^{22} we have the force on a single ion, viz. 5×10^{-10} mgrm.-wt. This force causes the ion to move with unit velocity. From these figures the reason is at once apparent why the forces are so enormous: their magnitude is due to the large number of atoms in one gram of hydrogen and the relatively large charge which is associated with each one of them.

In order to obtain a better idea of this charge let us suppose that we have a sphere of radius 1 cm. filled with water; the volume of water will therefore be $4\pi/3 = 4{\cdot}2$ cm.^{3}. Since the molecular weight of water is 18, it follows that 18 grm. or 18 cm.^{3} of water contain $60{\cdot}5 \times 10^{22}$ molecules; the sphere which we are considering therefore contains $14{\cdot}1 \times 10^{22}$

molecules. If now *only one* molecule from each million present loses its small charge, which is the charge carried by one monovalent ion, viz. $4{\cdot}77 \times 10^{-10}$ *E.S.U.*, then the total charge given up is $4{\cdot}77 \times 10^{-10} \times 14{\cdot}1 \times 10^{16} = 67 \times 10^{6}$ *E.S.U.* This would raise the voltage of the sphere by 67×10^{6} *E.S.U.*, which, expressed in volts, is 300 times as large, *i.e.* 2×10^{10} volts. This is a much higher voltage than could ever be produced artificially. It would correspond to a sparking distance of 6·7 kilometres if one is justified in extrapolating the results obtained with much smaller potentials. From this it is obvious that we shall never be able to deprive a body of its electric charge by gradually abstracting the charge from the body, for a condition would soon be reached for which the sparking distance was so great, that any further removal of the electricity would be impossible.

4. Galvanic Cells

The theory of galvanic cells is very intimately connected with that of the passage of electricity through solutions. Fundamentally, there are two kinds of galvanic cells, and although only one of them is of any importance when currents have to be produced, yet both are equally important to those who wish to understand the action of cells. The two types depend upon two essentially different phenomena; one on diffusion, and the other on solution. An explanation of the action of these cells in terms of the ionic theory was first given by **Nernst** about the year 1889.

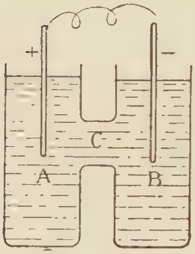


Fig. 132.

Concentration Cells. Let us assume that the two vessels A and B, Fig. 132, each contain sulphuric acid, but that the acid in A is considerably more dilute than that in B. The tube C makes connexion between the two acids. Because there will be fewer H^+ and $SO_4^=$ ions in A than in B, it follows that there will be an excess of ions tending to pass from B to A. The effective cause for this motion is to be found in the thermal agitation of the tiny particles which are present in solution; this agitation is also the cause of osmotic pressure if the motion is in any way checked [*cf.* p. 132]. If, however, the motion is not checked, then it causes diffusion to take place. Hence it follows that more particles will move from B to A than from A to B; the particles do not all move with the same velocity, however. The lighter H^+ ions travel more rapidly than the more massive $SO_4^=$ ions; this excess of hydrogen ions arriving in A causes the liquid there to become positively charged, whereas in B, which loses more H^+ ions than $SO_4^=$ ions, a negative charge will be acquired.

The positive charge in A will oppose the motion of H^+ ions into that region since like charges repel one another; on the other hand, the $SO_4^=$ ions will experience a greater attraction. This effect will be supported by the negative charge in B, for it will attract the H^+ ions towards B and urge the $SO_4^=$ ions towards A. These forces will soon become so great that there will be no excess of hydrogen ions travelling towards A; the charges in A and B will no longer increase. The process of diffusion itself, however, still continues, only the two kinds of ions will move with the same velocity and no further charges will be developed.

If now platinum wires are inserted in A and B, that which is immersed in A will acquire a positive charge with respect to that in B, and if the two wires are joined together by means of a conductor outside the circuit a discharge of electricity will take place from A to B. This will increase the rate at which the H^+ ions move, so that the charges on A and B will be renewed. A continuous current will therefore flow from A to B outside the liquid, whilst inside the liquid ions will be carried from B to A, and in so doing complete the electrical circuit. Does this process last for ever ? No, for the cause of the phenomena is due to the different acid concentrations in the two vessels and the passage of the current makes this difference become smaller and smaller. When the acid has diffused uniformly throughout the vessels the process comes to an end.

Replacement Cells. These cells have been used to produce electric currents for quite a long time. Let us assume that zinc sulphate solution has been placed in the vessel B, Fig. 132, and that it contains a rod of metallic zinc. Likewise A is filled with a solution of copper sulphate into which there dips a copper wire. This source of electrical energy is known as the Daniell cell, and it was described by **Daniell** as early as 1836. If the two electrodes are joined together the current passes from the copper to the zinc. Under these conditions it will soon be observed that the zinc is eaten away—zinc sulphate is formed. The chemical energy which is thereby liberated becomes converted into electrical energy. The explanation of the action of this cell was also given by **Nernst** in 1889.

Nernst set out from the fact that the metals of the anode in an electrolyte which pass into solution always possess an electric charge. For example, let us consider the electrolysis of copper sulphate. The SO_4 anion passes to the anode and there combines with an atom of metallic copper to form a new molecule of $CuSO_4$. But this process is intimately connected with the conduction of electricity through the solution, for the copper ions which pass into solution are those which carry the current across this section of the circuit. From this fact Nernst concluded that the metal atoms can, in general, only pass into solution when they exist as ions, *i.e.* when each atom is supplied with an electrical charge. Accordingly, he attributed to metals a certain solution pressure, for he assumed that metals have a tendency to pass into solution in the form of ions. If, therefore, a rod of zinc is placed in a vessel containing pure water, Fig. 133, the zinc will endeavour to dissolve in the water as zinc-ions. These ions will each carry a positive charge away from the zinc so that this will become charged negatively ; the water will become positively charged. The solution pressure varies from one metal to another ; it is greatest in the so-called base metals, *i.e.* these are the metals which have the greatest tendency to pass as ions into solution.

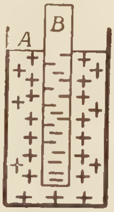


FIG. 133.

The objection will perhaps be raised that this assumption is quite untenable, since it asserts that zinc is soluble in pure water, a fact which opposes chemical doctrines. The answer to this objection must be made in two parts. In the first place an extraordinarily small quantity of dissolved substance is sufficient to create the small difference of potential which exists between the zinc rod and the solution [this small voltage can of course be observed and measured]. It will be remembered that in the numerical example on page 283 we were dealing with millions of volts. The potential difference of the element under discussion is of the order of one volt. But that substances dissolve in very small quantities when in contact with water is a view thoroughly familiar in chemistry. In the second place we have to reply that the potential difference due to the fact that the water is charged positively and the zinc negatively will oppose the passage of more zinc into solution, for the positively charged water will drive the zinc ions back whilst the negative charge on the zinc will hold them in check. After a very short time, therefore, the process of solution will cease ; this will happen when the forces due to the electrical charges which are produced are exactly equal and opposite to the solution pressure. If a number of zinc ions are already in solution, or if zinc sulphate is initially present in the solution, then the osmotic pressures will unite with the potential difference of the cell to oppose a further increase in ionic concentration.

If now a rod of copper is placed in the same vessel it will likewise endeavour to send copper ions into solution ; but since its solution pressure is smaller, it follows that the potential difference which exists when solution has stopped will be less than in the case of zinc. If, therefore, the rod of zinc has a potential $-1{\cdot}5$ volts with respect to the liquid, whilst that of the copper is $-0{\cdot}5$ volt, then if the vessel is connected to earth the potentials of the zinc and copper rods will be $-1{\cdot}5$ and $-0{\cdot}5$ volts respectively. Between the copper and zinc rods there will therefore be a potential difference of 1·0 volts, the copper being at a positive potential with respect to the zinc. When, therefore, electrical connexion is made between the copper and zinc outside the liquid, a current will flow along the external conductor from the copper to the zinc ; this current will pass through the solution from the zinc to the copper. It will reduce the potential difference between the electrodes so that the forces tending to prevent the solution of the zinc will be diminished ; more zinc will dissolve until the potential difference has again been created, or if the current is allowed to flow without interruption, the zinc will continue to dissolve until it has all disappeared, when, of course, the process terminates.

CHAPTER VI

IONS AND ELECTRONS

1. Cathode Rays

PRELIMINARY Facts. If one's finger is placed near to a rod of glass which has been rubbed, a slight crackling noise can be heard, and if the experiment is performed in a darkened room a very weak glow can be seen. The phenomenon can be made more clear by placing two spheres at a few millimetres distance from one another and charging them until a spark passes between them. If a Leyden jar is discharged, the noise which accompanies the passage of the spark becomes much more intense. This phenomenon can be much more easily observed with lower potential differences by placing two pieces of metal, which form the electrodes, opposite to one another in a vessel from which the air can be removed. When the air is gradually removed and the charges on the electrodes replenished by connecting them to some form of influence machine, the spark becomes a long band of light which soon broadens and fills the whole of the vessel. At lower pressures the band breaks up into striations which are separated by dark spaces. In short, such a series of phenomena are produced that one is simply amazed and hardly dares to hope for any explanation of them : and yet careful thinking has disentangled so great a part of this riddle, that it is difficult in these days for us to realize the utter helplessness of those who, fifty years ago, were confronted with this problem, and to appreciate the advances which have since been made.

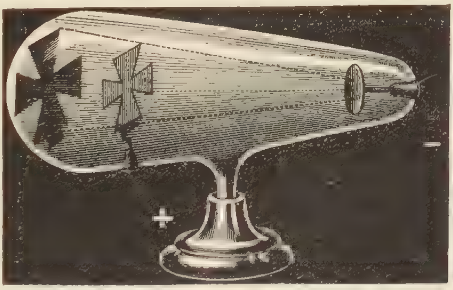


Fig. 134.

If the air is removed from a vessel until the pressure is about 0·01 mm., the phenomena become much more simple. The various moving striations and luminous columns disappear, and, instead, the walls of the glass vessel opposite the cathode exhibit a green fluorescence. Various researches, executed with great care, have shown that this fluorescence is caused by something emitted from the cathode which travels in straight lines to the walls of the vessel. This can be demonstrated by inserting an object between the cathode and the walls : if the object takes the form of a small metal cross, then a sharply defined shadow of it will be found on the walls (Fig. 134). If instead of hitting the glass this something is allowed to fall upon other fluorescent substances, various brilliant colours will be produced. This something is also capable of blackening a photographic plate.

If a pole of a magnet is brought near to the apparatus shown in Fig. 134 the shadow is deflected, a fact which proves that the path along which this something travels has become curved.

When an insulated plate is placed in the path it acquires a negative charge, so that one is forced to conclude that negatively charged particles are playing an important rôle in these experiments. Such moving particles will constitute an electric current; since the particles carry a negative charge it will be a negative current, *i.e.* the current will be directed towards the anode or away from the cathode. If the deflection produced by a magnet is considered in connexion with the rule concerning the deflection of a current by a magnet, given on page 256, then the above considerations will be found to be satisfied. These negatively charged particles which are sent out from the cathode along straight lines are known as **cathode rays.**

A considerable amount of energy is associated with these cathode rays. If the cathode is made in the shape of a concave mirror, then the rays are all brought together at a point not very far removed from the centre of the mirror, so that if a piece of metal foil is placed here it is soon raised to incandescence and finally punctured.

Our knowledge about cathode rays is largely due to the pioneer work of **Lenard.** In 1894 he showed that cathode rays were capable of penetrating a piece of aluminium foil not more than a few thousandths of a millimetre in thickness. In this way he was able to let the rays escape from the closed vessel in which they were excited through a small hole which was covered with a piece of the above foil. The foil was capable of withstanding the pressure difference between the air outside and that which remained inside the vessel, yet it was sufficiently thin to allow the rays to escape so that their properties could be further investigated. In air they no longer travel in straight lines but spread outwards in all directions; after traversing a distance of 2 or 3 cm. they can no longer be detected. But when the space outside the vessel, and in which the rays are being examined, is exhausted, then the rays are found to travel in straight lines. The greater the degree of exhaustion, the greater is the distance along which the rays can be detected and the more closely do their paths become rectilinear. If this outside region is filled with some gas other than air, but at the same pressure, then this distance becomes shorter when the molecular weight of the gas is increased. The distance does not depend upon the chemical properties of the gas. Cathode rays are only capable of penetrating a definite quantity of matter, after which they have lost their penetrating power—one imagines that they have also lost their velocity. The end of this distance can be detected in several ways: it is that point at which a fluorescent screen no longer glows, or a photographic plate is no more blackened, or an insulated metal no longer receives a charge.

The Nature of Cathode Rays. It is now necessary for us to endeavour to find out what we are really dealing with in the case of cathode rays. The large amount of energy which they transmit, their absorption in air and other gases, which depends only on the amount of matter through which the rays have passed, are facts which are closely connected with the view that we are dealing with material particles ejected with considerable speeds. It will perhaps be supposed that one could easily detect any matter which may be associated with these rays (as in the case of ions in electrolysis) by holding in their path a metallic screen. Or if they were a gas, then the pressure inside the space where they are would rise. All these attempts have led to no result, for **no one has ever been able to show that they carry with them any substance known to us.** On the other hand, however, they appear to be intimately dependent on matter, for if the amount of air in the discharge tube is further diminished the emission of rays suddenly ceases, and the passage of electricity through the tube cannot be obtained again even by increasing the potential difference across it. Attempts were therefore made to see whether these particles possessed anything in the nature of a molecular weight, or whether they carried a charge with them. Besides, a knowledge of the velocity of these particles would be of the greatest importance. Such measurements were first made at Cambridge by **Sir J. J. Thomson** in 1897.

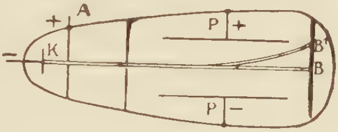


Fig. 135.

Let us assume that cathode rays are sent out from *K*, Fig. 135, in the direction *KB*, and that two plates are fixed one on either side of this path. If these plates are charged electrically, then any electrically charged particles passing into the space between them will be deflected. The amount of this deflection can be observed and measured by means of a fluorescent screen or photographic plate. Since the deflection by a magnet depends upon the charge carried by the rays and their velocity, this can also be employed to determine these quantities. This was actually done and we will now try to discern what these experiments tell us. We will assume that the strength of the field [*cf.* equation (99)] between the two plates is H_E. If the charge on each particle is ε, then the force urging them to one side is εH_E. If, further, the mass of each is m, the acceleration due to this force is $a = \frac{eH_E}{m}$. Each particle will therefore move with uniform acceleration in the direction of this force so that, in a time t, the distance traversed is $x = \frac{1}{2}at^2$ [*cf.* p. 22]. The time t during which they remain under the action of this force is the time in which they are travelling across the plates; it will therefore depend upon their velocity v, and, if l is the length of the path so traversed, $t = l/v$. If these values for t and a are inserted in the equation for x, we have

$$\frac{H_E l^2}{2x} = \frac{v^2}{(\varepsilon/m)} \quad . \quad . \quad . \quad . \quad . \quad (141)$$

In this equation l is the length of the deflecting plates whilst x is a length which is determined by the experiment; in fact it is the distance *BB′* which represents the displacement of the central spot on the photographic plate when the electric field is applied. H_E is the intensity of the electrical field, which is easily calculated from the equation $H_E = \frac{V_1 - V_2}{d}$, where $V_1 - V_2$ is the potential difference between the two plates, and d is their distance apart. The above equation still contains, however, three unknowns, or, if $\frac{\varepsilon}{m}$ is considered as one unknown for the time being, then there are only two, namely ε/m and v. This equation is therefore not sufficient to enable us to calculate them.

A second equation connecting these unknowns can be obtained by considering the deflection of a beam of cathode rays when a magnetic field is traversed. Electrically charged particles having a charge ε, and moving with a velocity v, constitute an electric current and as such will experience a deflection in passing through a magnetic field. The magnitude of the force which is operative on a such a particle has already been derived [*cf.* equation (127)] when we found $f = v\varepsilon H_M$, where H_M represents the strength of the magnetic field.

This force always acts in a direction perpendicular to that of the current. If we consider a north-seeking magnetic pole placed in front of the plane of the paper, whilst a south-seeking pole is behind this plane, the lines of force will be perpendicular to this plane, and the deflection of the current passing from K to B always occurs in the plane of the paper but perpendicular to the direction of the current at that point, exactly as with circular motion [*cf.* p. 42]. The particles will therefore be deflected either upwards or downwards according to the direction of the magnetic field through which they pass, but they will always be moving in a circular orbit. The magnetic forces will always cause the particles to move at right angles to their path ; the centrifugal force which thereby arises will tend to reduce the deflection. The curvature of the path will therefore be such that the centrifugal force is equal to force due to the applied magnetic field. If the mass is m, the velocity v and the radius of curvature r [*cf.* p. 43], the centrifugal force is mv^2/r. The circular arc which is thereby produced will be such that $v\varepsilon H_M = mv^2/r$, or

$$rH_M = \frac{v}{\varepsilon/m} \quad . \quad . \quad . \quad . \quad . \quad (142)$$

This is a second equation expressing a relationship between the same two unknowns v and ε/m. The radius r is easily measured, whilst H_M is obtained from a knowledge of the conditions producing it.

The results of these two experiments were as follows. The magnitude of the velocity was very large but it was not invariable ; in fact it was shown to be dependent upon the potential difference applied to the tube in which the cathode rays originated. It was discovered that the velocity of the rays was proportional to the square of the applied potential, a result which is most satisfactory in view of the following considerations. If V is the potential difference between the two electrodes and ε is the charge which moves across this distance, then the amount of work associated with this movement is εV [*cf.* equation (107)]. Further, if this work is used to accelerate a mass m, it must be equal to the kinetic energy produced. Hence

$$\varepsilon V = \tfrac{1}{2}mv^2 \quad . \quad . \quad . \quad . \quad . \quad (143)$$

so that

$$v = \sqrt{2V\varepsilon/m} \quad . \quad . \quad . \quad . \quad (144)$$

If, for example, the voltage varies between 2000 and 50,000 the velocity changes from 26,000 to 125,000 km. per sec. Or, if these velocities are compared with that of light, then the ratios are found to be 0·09 and 0·4 respectively.

The value of ε/m is of very great importance in discussing the nature of cathode rays. Actual measurements, however, have shown that this ratio is far more important than had ever been imagined. For example, it was found that ε/m had a constant value for rays which were not moving too quickly. At this point we may mention the fact that other processes were soon discovered, in which rays resembling cathode rays in all details could be produced. We shall learn more about these later on ; for the present it will suffice to know that for them the value of ε/m was always the same, providing that their velocity was not too great. From the fact that this value is continually recurring, we are forced to come to the conclusion that the ratio ε/m does not refer to some arbitrary mass carrying a charge which may be either large or small, but that it represents something far more fundamental. This becomes more significant since, as we have already seen, in dealing with cathode rays we do not find them to be associated with any kind of matter. In consequence of this, many attempts have been made to measure the value of ε/m with the greatest precision, and for that purpose all the various methods which offer any prospect of obtaining a reliable result have been utilized. The most reliable value hitherto obtained is

$$\frac{\varepsilon}{m} = 1{\cdot}77 \times 10^8 \quad . \quad . \quad . \quad . \quad (145)$$

when ε and m are expressed in coulombs and grams respectively.

In order to gain some idea of the importance of this number, let us consider once again how it has been obtained. The fact can no longer be doubted that the ratio ε/m refers to the individual cathode ray, and that it represents the ratio of the electrical charge to the mass of the particle. In ascertaining this value, however, it has not been necessary in any circumstances to determine the charge and mass of any single electron. The value of ε/m remains unchanged if both numerator and denominator are each multiplied by the same number. Hence it follows that we can find the value of this fraction by determining the charge and mass of any number of such particles, for the number so determined at once gives the ratio ε/m for each particle. When discussing the subject of electrolysis, we found that 96,495 coulombs of electricity were associated with one gram of hydrogen, so that $\varepsilon/m = 96{,}495$ coulombs per gram. If we compare these two values with each other, we find that the value (145) is greater than for the hydrogen ion. Other monovalent ions carry the same charge, but their mass is greater, so that for these heavier ions the ratio ε/m is smaller than for hydrogen; the value of ε/m is diminished, in fact, by the ratio of the masses of the two ions. For multivalent ions the electrical charge is increased several times, so that ε/m becomes greater accordingly.

When, therefore, we find that for cathode rays the ratio ε/m is 1800 times greater than for hydrogen ions, several possible explanations are open to us: either their mass is 1800 times less than that of the hydrogen ion, or else the charge on a single particle is 1800 times as large. It would also be possible that the actual state of affairs was intermediate between these two, for might not its mass be smaller and its charge correspondingly greater than that of the hydrogen ion?—*e.g.* if the mass were to be reduced 100 times whilst the charge were increased eighteen times the value of ε/m would be as before. The experiments which have so far been cited do not permit us to make any definite conclusion; but that the charge on the electron should be 1800 times greater than that on the hydrogen ion is a very improbable assumption, because so large a charge on such a small particle would give rise to repelling forces which were so great within it that the whole could scarcely be regarded as capable of existing.

But even if the charge on this particle is only a little less than 1800 times that on the hydrogen ion, it follows that **the carrier of this charge must be smaller than a hydrogen atom.** This is a very important result, for hitherto we have learnt to regard **the hydrogen atom as the smallest amount of substance which can ever occur in Nature.** Now we see that there are particles having a smaller mass. In fact, we have already seen that this small charge must be a fixed structural element of the world, since it never arises by chance, but is always found to be the same in magnitude no matter in what phenomenon it occurs. And if we should once find that the charge ε were exactly the same as that carried by an hydrogen ion, then the mass of a cathode ray would be the 1800th part of that of the hydrogen atom. Such knowledge would at once open to us new fields, and we should have obtained access to a quantity which was of an order of magnitude hitherto unknown. It follows, therefore, **that cathode rays are constituted from something which has so far found no place in our picture of the universe.** The moment this becomes known, a new era in research must begin. The frame in which our old picture was placed must be broken, for it no longer serves to encompass all our knowledge. What we therefore urgently require is some other relationship between ε and m, or, what is equivalent, we must try to measure one of them alone. We have no possible means whereby such a small mass as that of a single atom can be measured directly, but since the electrical forces are so large, we might hope that it would be possible to measure the charge on a single cathode ray particle.

2. THE DETERMINATION OF ε

From the beginning of this new era, physicists have realized that a determination of ε was most desirable and important. Many attempts have therefore been made to determine the magnitude of this fundamental charge ; they have been so numerous and diverse that it is impossible for us to describe them all.

The most reliable measurements were made by the American physicist **Millikan** in 1910. He succeeded in measuring the charge on a particle, the movements of which he was able to observe for several hours in succession. We will now discuss some of the details of his experiments. By allowing oil to issue through the jet of an atomizer he was able to obtain a very tiny spherical oil drop K, Fig. 136. Millikan also employed drops of mercury, and other substances. One such drop was allowed to fall through an opening O into a small vessel, the two ends of which were closed with glass plates, so that the drop could be observed by means of a microscope M when it was illuminated by a lamp L. The particle was observed to sink slowly and without acceleration just like the motes which can be seen in a sunbeam (these are very easily observed when a sunbeam is allowed to pass through a crevice into a room which is otherwise quite dark). For such small particles the dragging effect upon them due to the viscosity of the air soon becomes equal to the force which gravity exerts upon them; each particle then moves with uniform velocity v, which, for a particle of given volume, is proportional to its weight, *i.e.*,

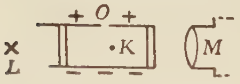


FIG. 136.

$$v = Cmg \quad . \quad . \quad . \quad . \quad . \tag{146}$$

where C is a constant ($g = 980$ cm./sec.^{2}). The upper and lower walls of the vessel consisted of metal plates insulated from one another. When these were electrified the particle was under the influence of an electric field the strength of which could be altered at will by the observer. If the particle carried an electric charge, then the electric force upon it was additional to that due to gravity. The effect of this additional force would of course depend upon the sign of the charge. The magnitude of this electrical force could be so chosen that the particle rose or fell, or that it should remain at the same spot for several hours on end; this latter condition arises when the electrical force is exactly equal and opposite to the gravitational force. When the particle did fall its velocity could very easily and accurately be determined. In the eye-piece of his microscope Millikan had two horizontal hairs, so that the time taken for the particle to move over a distance corresponding to that which separated these hairs could be determined with the aid of a stop-watch. The actual distance through which the particle had moved was found by calibrating the eye-piece; to do this a scale in millimetres is viewed under the same setting of the instrument. If H_E is the intensity of the electric field, whilst q is the charge carried by the particle, then qH_E measures the electric force upon that particle [*cf.* p. 289]. This must be added to the weight of the particle so that the velocity with which the particle moves is expressed by

$$v' = C(mg + qH_E) \quad . \quad . \quad . \quad . \quad (147)$$

Millikan allowed cathode rays, β-rays, or X-rays to pass through the instrument so that the charge on the particle could be altered; he then showed that the velocity changed spontaneously, whenever the charge was altered. The facts lead us at once to the following considerations. Let us assume that when the charge is q' the velocity is v''. We then have

$$v'' = C(mg + q'H_E).$$

Subtracting this from equation (147) we have

$$v' - v'' = CH_E(q - q'), \text{ or } v' - v = CH_Eq \quad . \quad . \quad (148)$$

If the charges q and q' can assume all arbitrary values, then their difference may be as small as we please; it therefore follows that the difference of their velocities $v' - v''$, or $v' - v$, become relatively just as small.

But if ε is the smallest charge or particle of electricity which can exist, then the difference $q - q'$ can only be due to the fact that a different number of such charges have become attached to the oil drop. The difference in velocity cannot therefore be made arbitrarily small. The velocity can only be charged by finite amounts. The smallest of such amounts will occur when the change in charge is equal to ε and is equal to

$$v' - v'' = CH\varepsilon \quad . \quad . \quad . \quad . \quad (149)$$

since $q - q' = \varepsilon$.

The change in velocity can never be less than this; it is true that it might be greater than this, but even so, its value cannot become arbitrarily large, but must always be an integral multiple of the above value. It therefore follows that for all observed values of v and v' there must exist some integer n such that

$$\frac{v' - v}{n} = CH\varepsilon \quad . \quad . \quad . \quad . \quad (150)$$

If the value of the constant C is known, then ε can be calculated, since all other quantities in this equation are known. The following table contains a few of the results obtained by Millikan for one particular drop, which he was able to observe for $4\frac{1}{2}$ hours; the electric field was such that the oil drop moved upwards and, having so moved, it was caused to fall again by merely reducing the field to zero. From all the observations made, only a few of which are recorded here, it was deduced that the velocity was always very nearly an integral multiple of 11·3. Considering the minuteness of the effects which were observed, this is a very remarkable result.

TABLE 27

v.	v'.	$v' - v$.	n.	$(v' - v)/n$.
− 42·7	+ 2·6	45·3	4	11·3
− 43·1	+ 25·5	68·6	6	11·4
− 43·1	+ 48·3	91·4	8	11·4
− 44·9	+ 70·4	115·3	10	11·5
− 43·9	+ 90·9	134·8	12	11·2
− 43·2	+ 115·6	158·8	14	11·3
− 43·8	+ 137·9	181·7	16	11·4

The first column contains the values of v, which is the velocity of the particle in the absence of an electric field, *i.e.* it is the velocity due to gravity alone. The second column records v', which is the value of the velocity upwards when the electric field is in operation, and the number of elementary charges on the particle is n. The last column contains the values of $(v' - v)/n$, and it indicates that the actual difference $v' - v$ is always an integral multiple of the number 11·3. The numbers do not specify the absolute velocity, since C had not been determined.

Millikan, however, was able to determine the value of C, which amongst other things depends upon the viscosity of the air. The mean value for the elementary charge ε as deduced from all his observations is

$$\varepsilon = 4{\cdot}774 \times 10^{-10}\ E.S.U. \quad . \quad . \quad . \quad (151)$$

This result is again very surprising, for it represents the same charge as we have previously found to be carried by the smallest ions in electrolysis. We therefore see that the smallest charge of electricity which can exist is an entity which we are justified in calling the **atom of electricity**, for it has all the properties which otherwise belong to a chemical atom. An electric charge smaller than this has never been found under any circumstances whatsoever; all charges which are greater than ε, however, are an integral multiple of this elementary charge (Millikan worked with oil drops carrying as many as 150 such charges).

By comparing the value of ε/m for cathode rays with that for H-ions, it follows that the large value in the case of cathode rays is to be attributed to their *smaller mass*, in fact, we are now quite sure that this mass is 1800 times smaller than that of an atom of hydrogen.

Even now one might still wonder whether or not an atom which had become attached to the atom of electricity had not also been caught by the oil drop: but the experiment is not vitiated thereby, since the mass of an atom with respect to that of any visible drop of oil is absolutely negligible.

3. THE ELECTRON

This particle of electricity which appears in cathode rays and which was measured exactly by Millikan and others, is called the **electron.** If these electrons do represent the atom and therefore the ultimate unit of electricity, it might be expected that they will be recognized in many other phenomena, since they must play an important rôle in all electrical processes. Even if all these processes are not equally suited as a means whereby the mass of the electron may be determined, yet it is important that we should be able to prove that we are actually dealing with these same electrons in all these phenomena, for herein lies the evidence that the electron is the ultimate and unique particle of electricity.

The Photoelectric Effect. When a newly cleaned or amalgamated piece of zinc Z, Fig. 137, which has been charged negatively, receive slight from an arc lamp B, it becomes discharged ; the plate not only loses its negative charge, but it even acquires a positive potential amounting to a few volts. If radiation continues to fall upon it no further effect is produced. A positively charged body becomes discharged when it is placed in front of such a zinc plate. Further investigations have shown that, under these conditions, electrons escape from the zinc plate ; they may be accelerated by means of an electric field. They are electrons, for they exhibit all the properties of a cathode ray beam : they are deflected by a magnet and by an electric field, and give the same value for ε/m. The number of electrons emitted is greater when the incident light is more intense. The final potential V_0 which the zinc plate eventually attains is very important in our quest for knowledge concerning this phenomenon. In particular, the effect of this positive potential is to prevent electrons from leaving the metal plate. So long as the potential is less than its limiting value, we see that the electrons will only be retarded ; they will still continue to leave the plate and thereby increase the positive charge on it. But when the potential has reached its limiting value, it is such that the electrons are just prevented from leaving the plate. From this fact we have to conclude that an electron is sent out from a plate which is uncharged with the same velocity as it would acquire in falling through a potential difference of $-V_0$. We are therefore able to calculate, from a knowledge of this final potential V_0, the velocity with which an electron is ejected from a plate whose charge is zero. It is given by [*cf.* equation (143), p. 290]

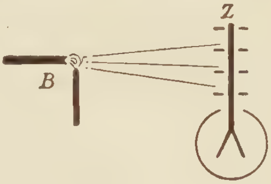


Fig. 137.

$$\tfrac{1}{2}mv^2 = \varepsilon V_0, \text{ or } v = \sqrt{2V_0\,\varepsilon/m}.$$

This velocity increases when the wave-length of the incident light becomes less. The so-called alkali metals—sodium, potassium, rubidium, and cæsium—emit electrons when light from the visible region of the spectrum is incident upon them. This fact has been utilized by **Elster** and **Geitel** to construct so-called photoelectric cells. By measuring the velocity with which these cells are discharged, a knowledge of the intensity of the incident radiation can be obtained ; this radiation may of course come from such a remote source as a star.

For us the most important fact which can be learnt from these various phenomena is that they confirm the electron's existence, for all their characteristics are those of cathode rays. In addition, it is found that the negative charge on a plate vanishes when it emits these same electrons. We might conclude that the charge, at least the negative charge, consists in general of a collection of electrons. Whether we are to regard a positive charge as a deficiency in the number of electrons normally present is a question which we shall proceed to discuss at once. Strictly speaking, we have only proved that a negative charge leaves the charged body in the form of electrons, and that another body on which they fall acquires a negative charge exactly equal in amount to the total charge carried by the electrons which it has captured. From these facts we have gathered something further, viz. that, in general, the charge on a body consists merely of a collection of these small electrical entities. The contrary assumption, that the electrons, when they are present in still greater numbers on a body, no longer continue to exist separately but collect together something like small drops of liquid to form a larger one, would be excluded as being untenable for two reasons: first, there are no facts to support it, and secondly, many quite unproved assumptions would also have to be included. Moreover, the electrons must repel one another, and therefore prevent any assemblage from taking place, because each one carries a relatively large charge. But that electrons, which in general, wherever they are met, appear as individuals having a definite size, charge and mass, should be fundamentally changed when they occur on bodies, is a hypothesis which would involve such radical changes in our ideas concerning the nature of electrons, that we dare not adopt it except on the evidence of well-established facts. Since such facts are not in our possession, we have to retain the idea that the charge on a negatively charged body is always only an accumulation of electrons.

Beta-rays. In a following chapter we shall become acquainted with β-rays, which are of radioactive origin, and learn that they too consist of electrons which are spontaneously sent out by the radioactive body. The β-rays carry the same charge ε as do the cathode rays. Their velocity is as great as the fastest cathode particle, and in fact approaches that of light itself. Their deflection in a magnetic or electric field, the fact that their motion is retarded as they move through air, and their other properties are likewise of the same order of magnitude and kind as we find in dealing with cathode rays. The proof that β-rays are electrons is thereby definitely established.

Thermions. In recent years another source of electrons has become very important. If the filament, F, in a completely evacuated electric bulb (Fig. 138) is raised to incandescence by means of a battery B, the wire emits a continuous supply of electrons. When another electrode is placed in the same lamp and raised to a positive potential, all these electrons are attracted towards it, so that a continuous stream of electrons passes through the apparatus from the filament to the anode, *i.e.* a current flows through the apparatus. But if the anode A is charged negatively the electrons are repelled and fail to reach this other electrode; no current flows in this opposite direction through the tube. This electric bulb possesses the very remarkable property that a current can only pass through it in one direction: no current will flow in the reverse direction since the conductivity of the tube in this direction is zero, there being nothing present by means of which the current can be carried.

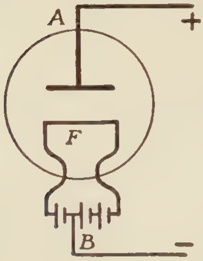


Fig. 138.

These tubes are widely used in industry. They convert an alternating current into a direct one since they prevent the passage of electricity through them in one direction. They are therefore **rectifiers.** By varying the strength of the filament current, the number of electrons emitted can be altered between wide limits. Moreover their velocity is at the disposal of the operator, as it is changed by merely altering the potential difference between the filament and the anode A. In an air-free space their velocity is given by equation (144). Since, in order to produce X-rays, it is necessary first of all to obtain cathode rays of definite velocity, these phenomena are utilized at the present time in X-ray work in the so-called **Coolidge** tubes.

Another important application of these tubes arises from the fact that electrons are deflected in a magnetic or electric field. It is therefore possible, by means of very small forces, to deviate an electron stream so that it impinges upon the anode at one point or another ; it can even be caused to fall first on one anode and then on a neighbouring one. Since the electrons have a mass which is inconceivably small, and because they move so quickly and without retardation, they are able to follow, for example, sound waves of the highest pitch. The tubes which are used in wireless telegraphy and telephony are designed after the manner shown in Fig. 139 ; between the incandescent filament F, which serves as the cathode, and the anode A a third electrode, G, is placed. This is called the grid on account of its grid-like appearance. Wireless waves, which may have a frequency of several millions, cause the charges on this grid to alternate between positive and negative with this same frequency. This causes the intensity of the stream of electrons to fluctuate with equal rapidity as it passes through the meshes in the grid. The variations in this current, which passes to the anode, are much more marked than those which were originally present in the wireless waves, which have been picked up and applied to the grid. It is for this reason that valves are much more employed to amplify small variations in current, than they are as rectifiers.

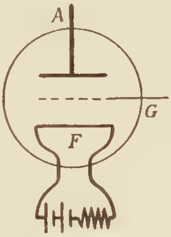


Fig. 139.

The Mass of the Electron. In speaking of the mass of an electron we have frequently had occasion to assert that the experimental conditions must be such that the velocity of the electron must not be too great [*cf.* the section dealing with the determination of ε/m]. We will now examine the relationship which exists between ε/m and v, for we shall find that this forms a very important aspect in connexion with our conception of the material world. At first the question will be considered quite free from every hypothesis and solely with reference to facts. Let us again consider the details of the experiment which led to equations (141) and (142), and from which the ratio ε/m and the velocity v could be ascertained. We shall now further assume that the value of ε has been shown experimentally [we refer especially to Millikan's work] to be invariable. We therefore assume that ε is a known quantity. Even so, there still remain two unknowns m and v in the two equations to which reference has already been made : these two unknowns can be deduced from experiment. We have already seen that it is quite easy to procure electrons travelling with different velocities. We shall imagine that this has been done and proceed to discuss the mass of the electron under these different conditions. The question is : Do electrons, moving with different velocities, have the same mass ? The numerous refinements which were made in the apparatus, necessary though they were for success, will not be dealt with here. The first experiments on these lines were made in 1901 by **W. Kaufmann,** who used the β-rays emitted from a certain radioactive preparation. He showed that the mass of an electron depended upon the velocity with which it was travelling, in fact, the mass increased with the velocity. It is therefore necessary to make a distinction between m_0, the mass of the electron when it is at rest, and m_v, the mass when the electron is moving with a velocity v. At a later date, **Buchner, Wolz,**

and **Neumann-Schäfer** made certain improvements in the details of the experiment, so that they were able to express the variation of m with v by means of an equation.

What has theory to say with reference to this ? In the first place, we know that in dealing with a stream of electrons we are really dealing with an electric current. And each variation in the strength of a current, in the above case each variation in the velocity of the electrons, requires the action of a force which alters the accompanying magnetic field when the velocity changes. On account of this performance of work the inertia of an electron appears to be greater when it is moving more rapidly. This change in inertia has been calculated by **Abraham**, and also by **Lorentz** and **Einstein**. These last two-mentioned research workers arrived at the same conclusion, whereas Abraham expressed the variation quite differently. The experiments, in so far as they permit us to make a numerical comparison, are all in favour of the result obtained by Lorentz and Einstein. Their theory expresses the mass m_v by means of the formula

$$m_v = \frac{m_0}{\sqrt{1 - \frac{v^2}{c^2}}} \quad . \quad . \quad . \quad . \quad (152)$$

where m_0 is the rest mass, and c is the velocity of light (3×10^{10} cm./sec.). This equation shows that m_v increases with v, and tends to become infinite when $v = c$. Since the mass is always determined by the force necessary to accelerate it, we may say that no finite force is sufficient to cause an electron to move with a velocity equal to that of light. This is also the meaning to be attached to the experiments represented by equations (141) and (142) : they were really an effort to determine the resistance experienced by an electron moving through electric and magnetic fields. From the numerous confirmations which the Lorentz-Einstein theory received, one began to suspect, in fact it was necessary to do so, that the mass of an electron was, in general, entirely electromagnetic.

Later on, however, equation (152) was extended by Einstein to all matter, whether it were electromagnetic or as mechanics would have us believe. In consequence of this, the above deductions have to be withdrawn. Einstein obtained these results by means of the theory of relativity, but his arguments are too difficult for us to repeat here. The result of his research was to show that no force is sufficient to make a body move with the velocity of light. It is not possible for us to verify this by the same type of experiment as was described for the case of a moving electron. Any velocity with which a body on this globe may move is not sufficient for a change in its mass to be perceived. We are quite certain, however, of one fact, and that is that no experiment contradicts the result obtained by Einstein.

We ought not to regard this change in mass as being due to **an increase in the number of atoms in a rapidly moving body, for this is most certainly not the case.** We determine the mass of a body as the resistance which it offers to any force tending to change its state of motion. According to Einstein's theory, this resistance becomes greater when the velocity of the moving body is greater, and it is in this sense alone that, according to the theory of relativity, the mass of a body varies with its velocity.

In the theory of relativity this relationship between mass and velocity is sometimes expressed in another way. Since c in equation (152) denotes the velocity of light, $\frac{v}{c}$ is for small values of v necessarily small and $\frac{v^2}{c^2}$ is of course still smaller. This enables us to put, in these circumstances, $1 + \frac{v^2}{2c^2}$ instead of $\frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}$. Equation (152) can therefore be written

$$m_v = m_0 + \frac{E}{c^2} \quad . \quad . \quad . \quad . \quad (153)$$

since $E = \frac{1}{2}m_0v^2$ is the kinetic energy of the mass m_0. The theory of relativity interprets this equation as follows: if a mass m_0 receives an amount of energy E, it increases its mass by an amount $\frac{E}{c^2}$, where c denotes the velocity of light. This statement can also be expressed as follows: **energy has inertia.** In fact, the inertia associated with an amount of energy E is E/c^2. Einstein himself maintains that this result, viz. that energy has inertia, is the most important and remarkable which the theory of relativity has so far yielded.

4. POSITIVE RAYS

At this point, the question may be asked as to whether or not there exist positively charged particles which correspond to the negatively charged electrons. With the answer to this problem there will also be found a reply to another, viz. are there one or two kinds of electricity? In the first place we will ask whether positively charged particles are found in a discharge tube similar to that shown in Fig. 138 [*cf.* p. 296]. The question was first raised by **Goldstein** in 1886. Let a series of holes be made in the cathode K of a discharge tube, Fig. 140, in order to permit any possible rays which may be passing from the anode A to the cathode K to be separated from the cathode rays, so that their characteristics may be investigated. If this is done, when the pressure on this side of the cathode is about 0·05 mm., the air will emit a feeble yellow glow: this is due to the so-called canal or positive rays which pass through the openings in the cathode into the region beyond. Since the days of Goldstein these rays have formed the subject of many investigations. In Fig. 140 these rays have their origin in the neighbourhood of the cathode itself; there are tubes, however, in which these rays originate at the anode. All such rays carry a positive charge; but this charge is not so indissolubly connected with them as in the case of cathode rays. In particular, it is found that during their motion they frequently change their charge so that they appear as carriers of a negative charge over a short distance; then they lose this charge and may again acquire a positive one. In 1898 **W. Wien** found that these rays were deflected by magnetic and electric fields: the deflection was opposite in direction to that found for cathode rays, a fact which proved that they were charged positively. He also showed, however, that this deflection was much smaller, from which it follows at once by

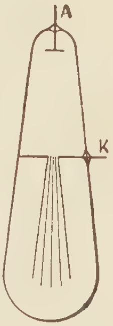


FIG. 140.

Reference Library

using the same formula (141) and (142) that the ratio ε/m is less for positive than for cathode rays. For a whole series of gases which were used to fill the discharge tube in turn, it was found that ε/m was of the same order of magnitude as in the case of electrolytes ; this establishes the fact that positive rays consist of charged atoms. The most accurate values for ε/m in the case of positive rays have been determined by **Aston,** who used his mass spectrograph, which we have already described [*cf.* p. 216]. Care was taken to see that all the particles in the positive rays had the same velocity, so that, by allowing them to move under the influence of electric and magnetic fields, he obtained deflections which depended only upon the mass of the particle. Hence, by measuring the deflection, he was able to calculate the mass. In this way he found that there were several kinds of chlorine, etc., facts which we have already discussed in the section on isotopes [*cf.* pp. 217 *et seq.*].

Moreover, in his experiments, **Millikan** had frequently noted changes in velocity which he attributed to the reception of a positive charge. In this way he was led to make observations concerning the smallest possible positive charge, although this was probably not included in his original programme. The results which he obtained, even in this respect, are of extreme importance, for his experiments showed that the magnitude of ε was independent of its sign.

At an earlier stage we have referred to the fact that there are two different views concerning the nature of the two opposite kinds of electricity. The first or dualistic hypothesis assumes that there are two essentially different sorts of electricity, which have to be associated in such a way that the opposite effects frequently discussed hitherto make their appearance. The second view is that there is only one sort of electricity spread uniformly throughout an uncharged body. If this electricity tends to collect in excess at one point, then that point acquires a positive charge, whereas a deficiency of electricity corresponds to a charge of the opposite sign. Hitherto it has been established with certainty that negative electricity exists as electrons, which are really a kind of atom or smallest particle of electricity. **If there is only one kind of electricity,** then this is certainly that which has arbitrarily been called negative : we shall see that this choice has not been a very propitious one. On this hypothesis a positive charge is merely due to a deficiency in the number of electrons present. Now if the dualistic hypothesis is adopted, then when one of Millikan's oil drops was found to have a smaller charge than previously, it meant that it had either lost an electron, or that it had captured a gaseous ion having a positive charge. If the magnitude of the positive atom of electricity were different from that of the electron, then this fact should have been detected experimentally : for the change which occurred in the velocity of the oil drop would have been equal to one of two values, according to whether a positive or negative atom of electricity had been responsible for this change in velocity. The experiments did not reveal any such differences. It therefore follows that if there are both positive and negative atoms of electricity, they are all equal in magnitude. Now it is extremely unlikely that, in general, two actual things should be so constituted that they always oppose each other in their actions : it is still more unlikely that even the ultimate constituents in these things should be exactly equal. But this difficulty is removed if one adopts the more simple assumption that there is only one kind of electricity.

X-rays. If this book were merely a history of the progress of science this would be the place at which mention of the rays discovered by **Röntgen** (1845–1923) in 1895 would have to be made. Very rarely has a discovery been so beneficial or promoted further researches than this one. The French physicist **Perrin,** who is certainly unbiased in his views on the subject, has said that the discovery of X-rays marked a new epoch in physics. X-rays, or Röntgen rays, are produced when cathode rays impinge upon a substance. It is quite a matter of common knowledge that these rays are able to penetrate bodies through which ordinary light cannot pass : it is owing to this fact that they have been so widely used in medicine. All this, however, lies beyond the scope of our book. But X-rays were, in a measure, rediscovered in 1912 by **Laue** and **Friedrich** and **Knipping,** who proved that they were really waves. In the following years, X-rays, in conjunction with light rays, have been investigated thoroughly, and at the same time proved themselves to be a very powerful tool as a means for further research. The detailed treatment of this subject will be postponed until we have dealt with light rays.

5. Gaseous Ions

From the foregoing arguments it will be quite easy to understand the following, even if, for the sake of brevity, all statements are no longer supported by detailed proofs.

In general, gases do not conduct electricity. Had it been otherwise it would neither have been possible to charge a body which was in contact with the air, nor for it to retain its charge for a considerable time. Actually it is found that no body will retain its charge indefinitely, and even **Coulomb** had noticed that the loss took place more rapidly than could be accounted for by faulty insulation. For if the sphere *K*, Fig. 141, is taken away, the insulating body (perhaps a piece of amber) will in no wise be changed thereby. Nevertheless one easily observes that an electrical charge on the electroscope remains there much longer. From this experiment Coulomb concluded that even the air is a slight conductor of electricity. But in 1882 **Giese** first recognized that this conductivity was due to the fact that the air had become ionized, *i.e.* the positively and negatively charged particles which had been created in the air were the agents by which the electricity was removed from the charged body—the process is very similar to the convexion currents of electrolysis. Later experiments have always confirmed this view, so that we are now able to assert with certainty the following :

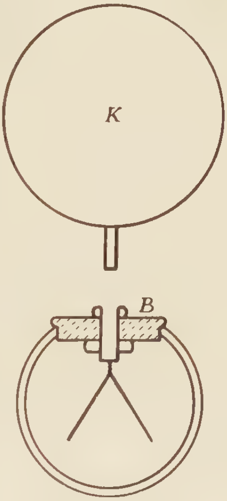


Fig. 141.

Gases, as such, are perfect insulators for electricity. Where they do appear as conductors, poor though they may be, this conductivity is always due to the ions present in them.

The Properties of Gaseous Ions are in several respects different from those of liquids.

1. Gases never become ionized spontaneously, whereas electrolytes are always ionized. They require some external agent, and we shall see that an agent is required which will constantly attack the individual

molecules of the gas, cause them to be shattered and to be bombarded. It has been shown that even the ever-present small conductivity of the air is due to the action of some such ionizing source.

2. With liquids it is only solutions of acids, bases, and salts which are decomposed by electrolysis; in fact, the decomposition is always such that the following rule is valid: the hydrogen of acids and the metal of the salt are always charged positively, whilst the remaining acid radicle or the OH-ion of a base is always negative. Moreover, the valency of the charged particle plays a decisive rôle. With gases this is not so. All gases can be ionized—even gases such as O_2 and N_2 which only consist of one element, and monatomic gases such as He, A, Ne are capable of being ionized.

At least for these last-mentioned gases, the ionization can only arise when an atom has acquired or given up a charge, *i.e.* when it has received a negative or a positive charge. For such examples as these, when the rule which applies states that positive and negative electricity must always be produced in equal amounts, it follows that the number of positively charged helium atoms must always equal the number of negatively charged particles corresponding to them. The process can be visualized as follows: first an electron is removed from some atom, which thereby becomes positively charged. This free electron is attracted by an uncharged atom, unites with it, and so produces a negative ion.

3. Gaseous ions are still more independent of one another than those in liquids. In an electrolyte the total positive charge is equal to the total negative charge, so that, as a whole, the liquid appears to be uncharged; this is not necessarily true for gases. It is still true that the production of a negative charge is always associated with the creation of an equal positive one, but after a little while these two charges can be at great distances from one another. When an ionized gas is allowed to pass by a positively charged body the negative particles are attracted to it so that the body becomes discharged. The positive ions take no part in bringing about this discharge, and the gas, as a whole, moves past with a positive charge. If this gas is compressed into some vessel, then this latter receives a positive charge.

4. The mass of a gaseous ion is not previously fixed as with those of a liquid. Even after the ion has been formed it can attract other atoms, so that the total mass of the ion may become considerable. In damp air, in particular, the ions become associated with so much water vapour that their motion is very considerably impeded. As a general rule negative ions are somewhat lighter than positive ones.

5. Whereas electrolytes remain ionized continuously, the ionization of a gas is lost, perhaps in a few minutes, after the ionizing influence has been removed. This can only be due to a recombination taking place between ions of opposite sign. The electrical forces will cause them to approach one another, so that after they have united the gas appears, as a whole, to be uncharged, and the normal gaseous state has been reached. The probability that an encounter will take place between *some specified positive* ion and a negative one will be greater when the number of negative ions per unit volume is increased. The probability that any one positive ion shall combine with a negative one also increases as the number of positive ions present increases. Since now, in general, that is when no special influence has produced an excess of ions of any one kind, just as many positive ions as negative ions are present, the number of ions N

which combine together per second will be proportional to n^2, where n is the number of ions of one sign still present. Hence

$$N = Cn^2 \quad . \quad . \quad . \quad . \quad . \quad (154)$$

where C is called the coefficient of recombination.

6. When the ions are subject to an electric field, a condition which may be realized by placing two electrodes at different potentials in the gas, then the ions will all be attracted to these electrodes ; the particular one to which they pass will, of course, depend upon the sign of their charge. If, at the same time, the ionizing agent continues to exert its influence upon the gas, a continuous current will pass between the plates. But the law of this current is not the same as for currents in electrolytes. Initially the intensity of the current will depend upon that of the ionizing effect, *i.e.* upon the number of ions produced per second. When we consider the effect of the applied voltage on this current, we have to remember that there are two influences at work tending to diminish the number of ions present : first, the potential difference itself, which removes the ions to the electrodes, where they lose their charge; and secondly, the recombination which takes place amongst the ions. Each influences the other to some extent. The greater the force due to the electric field, the more quickly do the ions move, so that the chance of a recombination taking place before they reach the electrodes becomes less. The manner in which the current varies with the applied potential is therefore somewhat as follows :

By increasing the potential a stage is easily reached at which the ions move so quickly that the amount of recombination which occurs is negligible. Under these conditions, the number of ions passing to the electrodes is equal to the number produced by the ionizing agent. Hence, by measuring the strength of this current, the intensity of the ionizing agent is ascertained, whatever may be its constitution. When, therefore, the potential is increased still further, the current can no longer increase, since all the available ions have already been utilized. The current is therefore not represented by Ohm's law, for it no longer increases with the potential difference. The current is said to have reached its **saturation** value. If the potential is smaller, then the velocity with which the ions are removed is reduced, so that the current becomes less : the reasons for this are twofold. In the first place the ions move more slowly, and secondly, some of the ions will be able to recombine as they journey through the gas. The current will therefore diminish more rapidly than can be accounted for by Ohm's law. The intensity of the current for various potential differences is represented in Fig. 142. The shape of the curve is the same for all processes in which a definite and limited number of ions are produced, and amongst which recombination can take place.

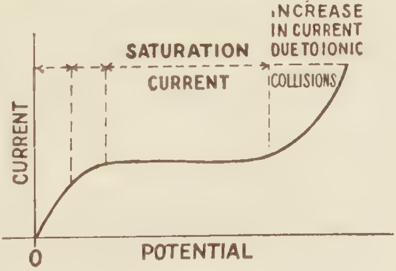


Fig. 142.

When *no* potential difference exists between the two plates the number of ions per unit volume will also be constant. Since the number of ions disappearing owing to recombination depends upon the number present, it follows that the number of ions formed under the action of a constant ionizing agent will at first increase ; this causes an increase in the rate of recombination amongst the ions, until finally a state of equilibrium is reached in which the number of ions disappearing owing to recombination is equal to the number created in the same time.

If now a very small potential difference is applied to the electrodes the number of ions conveyed with the current is so small that the total number present is not appreciably changed. But since the velocity increases with the potential the initial variation in the current is represented by Ohm's law. The curve, Fig. 142, is therefore at first a straight line which rises according to Ohm's law. Finally, when the current has attained its saturation value the curve no longer rises ; at points in between the slope of the curve gradually passes from one value to another. At very high potentials the curve again slopes upward ; an explanation of this effect will be given shortly [*cf.* p. 306].

Ionizing Agents. There are a great number of ionizing agents known, which are capable of producing ionization in a gas. They are as follows : high temperatures, chemical processes, collisions between molecules and rapidly moving positive and negative particles (ions and electrons), and finally light rays, including X-rays.

High Temperatures. Metals, when raised to a temperature of several hundred degrees, begin to ionize the air around them. Initially the ions are all positive, but at temperatures higher than about 900° *C.* positive and negative ions appear in equal quantities. It is certain that even at the beginning the number of positive ions must be accompanied by a production of an equal number of negative ones, but these do not leave the metal. It is supposed that the absorbed gases are expelled from the metal when the temperature is raised, and thereby lose an electron. Why it is lost we cannot say. Possibly it is due to a collision with another particle. Even if, in general, the energy of the collision is not sufficient to cause an electron to be ejected from an atom, yet some few particles will always possess an amount of energy much greater than the mean, so that they will be able to produce ionization.

Flames. In flames we have not only high temperatures but also chemical processes at work, during which large amounts of energy are exchanged. Flames are always ionized to a large extent, ions of both signs being present. This can easily be shown by bringing a charged body so near to the flame that the ions become attracted by it. Ions of sign opposite to the charge on the body will pass to it, and the body eventually becomes discharged. This always takes place irrespective of whether the body is charged positively or negatively. When salts, especially those of the alkali metals, are introduced into the flame, the ionization is considerably increased.

Chemical Processes. When a piece of phosphorus is allowed to burn the surrounding air becomes ionized. These ions, however, are associated with so much water that they are very heavy and move slowly. In consequence of this fact recombination takes place less rapidly than with other ions.

Electron Impact. Still more important is the ionization produced by other ionizing agents, especially corpuscular rays. All corpuscular rays are capable of producing ionization when their velocity is sufficiently high, *i.e.* they cause an electron to be ejected from its parent element. Since electrons always carry a negative charge and the atom, as a whole, is uncharged, it follows that the part of an atom left when an electron has been ejected must be charged positively. From this we may conclude that an electron is bound to the rest of the atom by a certain and definite force which must be overcome by the ionizing agent. Almost involuntarily we come to regard the release of an electron from an atom as the result of mechanical actions produced by the corpuscular rays which, whilst travelling very rapidly, have passed into the atom itself. Again and again this conception has been shown to be reliable. Since the velocity of cathode rays can be fixed at will by causing them to pass through a region where a definite potential difference exists, it is possible to ascertain exactly how great their velocity must be in order for them to cause ionization in gases. In this way it has been shown that this velocity varies with the gas to be ionized : this implies that the force with which an electron is attached to the rest of the atom varies with the nature of the gas. The potential difference necessary to impart a certain velocity to an electron is given by equation (143), so that it is usual to express the velocity of an electron in volts. Hence, when we say that the ionization potential of helium is 24·5 volts, we mean that cathode rays must pass in vacuo (so that the effect of collisions with other molecules is negligible) across a potential difference of 24·5 volts in order for them to acquire sufficient velocity to ionize helium atoms. The first determinations of ionization potentials were made by **Franck** and **Hertz** in 1913.

The values of the ionization potentials for a few gases are recorded in Table 28.

TABLE 28

Ionization Potentials

Element.	Volts.	Element.	Volts.
He	24·5	O_2 . . .	15·4
Ne . . .	21·5	Hg . . .	10·38
Ar	15·4	Na . . .	5·13
K	13·3	K . . .	4·13
X	11·5	Rb . . .	4·16
H	13·5	Cs . . .	3·96
H_2	15·7		

From a knowledge of ionization potentials many important conclusions have been made concerning the nature of the binding between electrons and the rest of the atom. We shall soon have to return to a detailed discussion of this subject.

Ionic Collisions. Like electrons, ions themselves can produce ionization when their velocity has become sufficiently large. In the next chapter we shall become acquainted with rays sent out by radioactive bodies and known as alpha-rays. They carry positive charges, and their velocity is so great that in vacuo even the slowest of them would travel, in one second, a distance equal to that from the north pole to the equator ; the fastest would travel twice as far under similar conditions. They do not travel so far in air, for they collide with air molecules and in so doing produce an enormous number of ions. This number for alpha particles varies from 100,000 to 200,000 pairs according to the velocity. In consequence of this, their velocity gradually diminishes ; the slowest of them do not travel in air more than 3 cm., whilst the fastest only travel 8 cm. Their velocity is then about $0{\cdot}82 \times 10^9$ cm. per sec., which is approximately one-half their initial velocity. Having reached this stage, they are no longer capable of producing ionization; their retardation is increased, and the rest of their energy is soon converted into heat, after which they are at rest.

From the remarks made, it is almost self-evident that the positive and negative ions which are produced in the air are capable of creating further ions if their velocity is sufficiently great, *i.e.* an electron will be hurled forth from any gas molecule with which the ions may collide. If these happen to be subject to an electric field, new ions will be produced, so that the number of ions present increases very rapidly; the conductivity will attain those high values denoted by the right-hand portion of the curve shown in Fig. 142 [*cf.* p. 303].

If the charges on the electrodes are replenished continuously by some electrical contrivance, so that the potential difference between them remains constant, then luminosity may be produced. We have examples of this in the arc lamp itself, and in **Geissler** or similar tubes where the discharge is always accompanied by the production of light. On other occasions, as when an influence machine is used, the discharge ceases until the requisite potential difference has been regained. The discharge is intermittent and is known as a spark discharge.

Light Waves. Even light waves can produce ionization. The strongly electropositive alkali metals, sodium and potassium, which very easily give up an electron, emit electrons when visible light is incident upon them. For zinc and other metals it is only violet and ultraviolet light which is effective in this respect. X-rays which are not too soft, *i.e.* which do not approach the ultraviolet waves too closely, are capable of removing electrons from all elements, even gases, and thereby produce ionization in them. Gamma rays, which are ejected by radioactive bodies, act in a like capacity. At a later stage [*cf.* p. 339] we shall discuss the supposed mechanism whereby waves of light are able to set electrons free.

6. THE PASSAGE OF MATERIAL RAYS THROUGH SOLID SUBSTANCES

One might be inclined to believe that the passage of material rays through substances would only be of little importance to us; actually, the exact opposite is true. Let us suppose that we had a body with whose internal structure we wished to become acquainted, but we were not able to open it. Eventually we might have found that it was possible to insert very fine long needles into this body. Then we might at least ascertain whether the body were uniformly soft, or whether hard inclusions were embedded in it; in this latter case their size and number could be estimated. Such fine needles which can pass right into a body, and by means of which we are able to discern the nature of its innermost parts, are the moving ions; the electrons are still more fine than these.

In order to obtain this knowledge we have to shoot these particles with considerable velocities into the body. Let us now see what happens to such a projectile when it gets inside the body. The first attempts in this direction were made as far back as 1902, when **Lenard** investigated the paths of cathode rays which had passed through thin metallic foil. He discovered a very remarkable result. The effects produced with individual rays varied between wide limits; some passed through quite uninfluenced by the metal, just as if they had travelled through a vacuum;

others were deviated from their path several degrees so that the pencil of rays which had been sharply defined before entering the metal became diffuse, whilst a few individual particles were deflected sidewards, some even almost retracing their original path. From these observations Lenard concluded that inside an atom there were points around which the field of force was very concentrated ; the extent of this concentrated field is very small, even in comparison to the distance between two neighbouring molecules, whilst the space between these centres of force, like that between the atoms themselves, is empty. Soon we shall become acquainted with other facts which lend support to this hypothesis ; when this stage in our progress is reached, the discussion will be renewed.

Resumé. Electricity, like the elements of the periodic system, is built up of atoms. The charge on an atom of electricity, or the elementary charge, amounts to $4{\cdot}77 \times 10^{-10}$ *E.S.U.* All charges residing on bodies, of whatever kind they may be, consist of a more or less large number of these small particles of electricity.

The negative charge, when it exists as a free electron, is without mass (in the usual meaning of this word as applied to an element belonging to the periodic system), but it possesses inertia. The inertia of an electron is 1800 times smaller than that of the hydrogen atom, and is the smallest material particle which has hitherto been isolated.

The electron is to be found in all atoms, even though they are uncharged ; it is a constituent of all atoms and, as such, forms the smallest structural element of which the material world is built up.

When an electron leaves a body, this latter acquires a positive charge. Positive charges have only been observed in combination with atoms or molecules of the periodic system. Even when an atom possesses equal positive and negative charges so that it appears to be uncharged, yet in the atom itself these two charges are independent and separated from one another.

CHAPTER VII

RADIOACTIVITY

1. The Discovery of Radium

INSPIRED by the fact that Röntgen had discovered rays capable of penetrating opaque bodies, **Becquerel** (Paris) imagined that fluorescent minerals sent out similar rays. He actually found that a photographic plate was blackened by their action. This effect was not prevented by the black paper in which the plate was wrapped. By investigating all possible minerals it was finally decided that uranium was the element responsible for this phenomenon. Soon, however, **Mme. Curie** discovered that there were other substances which exhibited these characteristics much more strongly than did uranium. If the effect were really due to uranium, then it should be greatest in pure uranium. Since this was not so, however, Mme. Curie concluded that the effect was to be attributed to some other substance with which the uranium was always associated, but the quantity of this substance was so small that the usual methods of chemical analysis had failed to detect it. If it could be isolated, the rays from it should be very intense. She resolved to collect some of this new substance by abstracting it from a large mass of uranium ore, notably pitchblende, and then examine its properties; this plan, in which she was assisted by her husband, **Pierre Curie**, was duly carried out.

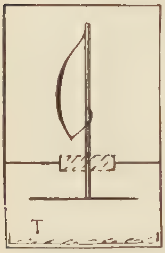


Fig. 143.

The Detectors of Radioactive Substances. Next to the photographic plate, the instrument which proved itself to be most useful in detecting these new substances was the electrometer. In particular, it showed that the rays from these bodies, just like Röntgen rays, were capable of ionizing the air in their immediate vicinity. A charged electroscope, which was otherwise well insulated, lost its charge very rapidly when it was placed near to a radioactive substance; by measuring the rate at which this charge was lost, one had a very convenient, quick and sure method of comparing the intensity of the effects produced. If one places the same amount of substance on the plate T of an electrometer, Fig. 143, and observes how many divisions on its scale are traversed by the moving part per unit time, then the strength of the radium preparation is greatest when the motion is most rapid, always provided, of course, that the potential difference is so high that the current through the instrument has reached its saturation value [*cf.* p. 303]. If $V_1 - V_2$ is the drop in potential per second, whilst C is the capacity of the instrument, then the corresponding loss in charge is $C(V_1 - V_2)$, and if the current has its saturation value, then this is the total charge carried by all the ions of one sign which are produced in one second. Since each ion has a charge ε, where $\varepsilon = 4{\cdot}77 \times 10^{-10}$ *E.S.U.*, the number of ions present is obtained at once by simple division. When we are dealing with radioactive preparations of the same type, this number will be proportional to the quantity of radioactive substance present. If such measurements are to be used to estimate the absolute amount of radium present, then due attention must be paid to the following points : Are all the ions produced actually counted by this method, and does this number always correspond to the amount of radium, and so on ?

A third means was soon found whereby radioactive substances could be investigated. Fluorescent materials emit a pale light when they are brought near to an active preparation. The scintillations which are produced on a zinc sulphide screen *S*, Fig. 144, when it is placed close to a radioactive body *R*, may be observed with the aid of a magnifying glass *L*. These scintillations have been both a means of investigating radioactive bodies and of enabling us to demonstrate radioactive effects very simply. The electrometric method is the most important of the three, because it enables us to say at once and with certainty what is the strength of the effect. This instrument was of constant service to the Curies in their attempts to produce pure radium from minerals.

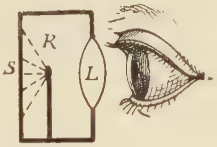


Fig. 144.

The Method of Isolating Radium. The process which ultimately met with success was essentially as follows. When they extracted from the natural ore by various physical and chemical means the substances which were contained in it, they found a whole series of substances which did not exhibit radioactive phenomena. For example, when the crushed ore was agitated with sulphuric acid, one part dissolved whilst the other did not. Both portions were then examined by the electrometer, which showed that the portion which did not dissolve was radioactive. In this way, therefore, it was possible to obtain a relatively large amount of active material in a somewhat small space. Finally, two products were obtained from pitchblende which resembled barium and bismuth, but which were able to discharge an electroscope or blacken a photographic plate more than a hundred times more rapidly than pure uranium. The Curies concluded (1898) that the radioactive substances associated with the barium and bismuth were exceptionally active. These radioactive substances must be very similar to the substances with which they were associated, since nobody had previously been able to separate them. After a year of tedious labour it was shown that the radium content of the barium and the polonium content of the bismuth had been increased many times.

The most important thing about these discoveries is not that a few new substances were revealed, but that new facts concerning the elements of the periodic system were made known. Until then these facts had been quite unknown, yet they have since enabled us to increase and deepen our knowledge of the elements very considerably.

In the next three sections we shall discuss first the rays emitted by radioactive bodies, then the radioactive bodies themselves, and finally the deductions concerning the structure of the atom to which they have led.

2. The Rays from Radioactive Materials

Even the first investigations concerning the rays sent out by radioactive substances produced such a multitude of new facts that much confusion was caused until **Sir Ernest Rutherford** succeeded in co-ordinating them.

If a radioactive compound is placed inside the instrument depicted in Fig. 143, and its discharging action is examined, it will soon be found that the intensity of this action is reduced about a hundred-fold when the preparation is covered with a piece of thin tinfoil. A second piece of foil does not make any appreciable difference, but if a piece of lead about 1 mm. thick is used to cover the material, the discharging effect is again reduced about a hundred-fold. Even after this, there still remains an effect which is clearly perceptible, and this can only be reduced to a negligible amount by using a cover of lead several centimetres thick. From these facts Rutherford concluded that such a radioactive material emits three kinds of rays, which he called the α-, β- and γ-rays. The penetrating power of the alpha rays was least, whilst that of the gamma rays was greatest.

Since at this time cathode rays were known and had been shown to consist of electrically charged particles, attempts were soon made to see whether or not these newly discovered rays were similarly charged. According to the law of action and reaction, if they are charged they will be accelerated under the influence of electric and magnetic fields ; or, what is more expedient for purposes of observation, they will be deflected from their path. Let us suppose that a radioactive preparation is covered with a piece of thick lead in which a small hole has been drilled, Fig. 145. If a magnetic field, whose lines of force pass perpendicularly from the front to the back of the paper, is applied, then those particles in the rays which are positively charged will be deflected towards the left, whilst those having a negative charge will experience a deflection towards the right.

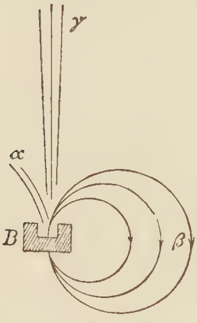


FIG. 145.

The result of this experiment was to show that a portion of the rays was deviated towards the left, another towards the right, whilst quite a definite portion was not deflected at all. When the radium preparation was covered with a piece of aluminium foil the positive pencil of rays vanished ; similarly by using a piece of lead 1 mm. thick the negative pencil was made to disappear, whereas the uncharged portion was only absorbed by a layer of lead several centimetres in thickness. In this way it was shown that the α-particles are charged positively, whereas the β-rays carry negative charges, and the γ-rays are not charged at all.

From a measure of the deflections produced, the velocities of the particles could be estimated as already done for cathode rays. The α-rays have a velocity equal to about $\frac{1}{15}$ that of light ; β-rays travel much more quickly, the fastest moving with a velocity of less than 2 per cent. below that of light, whilst the velocity of the slowest is about one-third that of light. The velocity of the γ-rays could not be calculated from these experiments, since they were not deflected. At this point we may anticipate the results which have been deduced from other sources and state that the velocity of γ-rays is equal to that of light. The γ-rays are really æthereal waves, and only differ from visible light by the fact that their wave-length is much shorter. Their wave-length is not only the shortest of all wave-lengths (cosmic rays excepted), but it is the shortest distance which we have been able to measure with certainty. Its order of magnitude is one thousand-millionth part of a millimetre.

All three types of rays have their intensity reduced when they pass through air. The reduction is greatest for the α-rays, which have only a range of 3 to 8 cm. in air. Whilst they are moving across this distance they ionize the air very considerably, a fact which is due to their frequent encounters with the air molecules. The end of this range is that point beyond which the rays are no longer able to produce ionization or to blacken a photographic plate. By their velocity, negative charge, and mass, β-rays prove themselves to be merely electrons, but they are electrons with a very high velocity. They are able to travel much farther in air than the α-rays, but the ionization which they produce is only about one-hundredth of that produced by an α-ray travelling through the same distance.

The penetrating power of γ-rays is much more marked, for, like X-rays, they are able to pass through thick sheets of lead. When they have passed through 100 metres of air at normal pressure, their intensity is only reduced to one-half its initial value. Accordingly, the number of ions produced by γ-rays is about 100 times less than that due to β-rays passing across the same distance.

Owing to the newness of this type of phenomenon, when these rays were discovered, all attempts to account for them were attended by considerable difficulties. Doubts were held even concerning the source from which these rays derived their energy. The following facts, however, played a very decisive rôle in the development of a theory which would explain radioactive phenomena.

1. If a definite quantity of radioactive material emits a certain amount of radiation, then this radiation is not changed even when the substance enters into a chemical reaction. Whether the radium is present as an element, or whether it exists as a chloride, bromide or sulphate, or in combination with other materials so that a new molecule is produced, the ionization which it causes in the air is always the same.

2. If the substance is then investigated at different temperatures ranging from that of liquid air to that of an electric arc, the radiation is likewise unaltered. Now all chemical processes depend very largely on the temperature. At the temperature when flame is produced hydrogen and oxygen combine with a violent explosion to form water, whereas at room temperatures no combination takes place between them. On the other hand, when water is introduced into an electric arc it is dissociated into its constituents. The velocity with which a chemical action takes place can be altered at will by merely changing the temperature, but we are quite powerless to vary the radiation emitted by a radioactive body. We can only stand idly by and observe these changes taking place irrespectively of all external conditions. At the present time about thirty different radioactive substances are known and, without exception, external influences have no effect upon them. Since these processes are independent of other elements with which the radioactive elements may be combined, it is certain that the radiation is not to be attributed to some action taking place between the active and some other element combined together to form a molecule, but that we are here dealing with a process in the active atom itself. This is confirmed by the second fact, viz. that the radiation is independent of the temperature, a criterion which does not apply to other chemical reactions.

3. This same hypothesis is confirmed by yet another phenomenon. All radioactive materials have a temperature which is always a little higher than that of its surroundings, *i.e.* heat is continuously being produced. In preparations containing a large amount of active material this thermal

energy can be measured, and it is then found to be almost a million times larger than that liberated by the most violent of chemical reactions, provided, of course, that the same mass of material is always contemplated. The heat changes which occur are therefore of quite another order of magnitude to those which relate to the more usual chemical changes. The heat which here appears suddenly cannot be produced by chemical change ; we have to assume that we are dealing with a new source of heat in the element itself. That no previous observation of this large amount of heat had been made is simply due to the fact that radioactive materials are only found in Nature in a state of almost infinite dilution with other substances ; as soon as a large enough quantity had been collected the heat energy was observed.

By means of these rays, it was soon discovered that quantities of material could be used for experimental purposes, which were so small that the most sensitive chemical balance would never have been able to detect them. In this way our knowledge of radioactive substances increased very rapidly.

3. RADIOACTIVE SUBSTANCES

In addition to the remarkable radiations which these substances had been shown to emit, still more surprising results were soon discovered in connexion with the substances themselves. Even in 1899, the Curies had noticed, in the course of their work, that all bodies which had been in contact with radioactive matter for a little while acquired the property of emitting radiations themselves. This was called the ‘ induced activity ’, a term which is not to be used to mean more than the observed facts. The induced activity was due to a material deposit which could be removed by rubbing or, like a film of water, by heating it ; and, like water, it condensed again on a cool surface.

A year later several research workers discovered simultaneously that a gas was continuously being developed from various radioactive materials, and this gas showed all the properties characteristic of such substances. These gases were called emanations, and it was further found that the induced activity was always associated with this gas. Whenever it was absent, there was no induced activity ; if a gas-tight cover prevented the gas from escaping from the parent substance no induced activity was found outside the cover. But wherever the gas could penetrate, there the induced activity was also found.

A new problem then appeared. The gas which had been produced gradually vanished. If one had initially a large gas-holder containing air rich in emanation and removed the same amount daily in order to test its activity, then each day the activity of the specimen examined was found to be less. After 3·8 days the activity was reduced to half its initial value ; after the lapse of another 3·8 days the remaining activity had not vanished but had only been reduced to one-half, *i.e.* to one-quarter its initial value. And so the process continued. After each 3·8 days the activity was reduced by one-half. A whole series of other radioactive substances were found to behave in the same way, but the magnitude of the half-value period had for each its own particular value, which varied from a fraction of a second to several thousand millions years.

The Theory of Atomic Disintegration. Supported by these and other facts, **Rutherford** and **Soddy** put forward a hypothesis, in 1902, which was a suitable explanation of all the phenomena then known. However impossible this hypothesis appeared at that time, and however deeply it changed the ideas, which were then prevalent, concerning the innermost nature of bodies, it was very soon confirmed by a whole series of new facts, so that its validity can no longer be doubted. The hypothesis assumed that atoms could disintegrate, or more correctly, that the explosion of an atom was accompanied by the emission of radiation. Their line of thought was somewhat as follows.

The emission of rays from radioactive bodies is a process which takes place in the individual atoms alone: nothing else is a participant in it. We see that the radiation is accompanied by the ejection of material particles from the body, even from the atoms themselves. In consequence of this loss of matter the nature of these bodies must of necessity be changing. What remains after the atom has sent out such a ray is quite a new body, with a lower atomic weight, and consequently with different physical and chemical properties—it is a new element, but, like all other elements, it belongs to our periodic system of elements and must find a place therein. When an atom of radium emits an alpha particle the remainder of the original atom forms an atom of the emanation. This emanation is a gas, an inert gas, which has its place in the periodic system along with the other inert gases—in fact, a place had been left in the table for such a gas [*cf.* Table 5, p. 111]. In 3·83 days the activity of the emanation was reduced to one-half; a new substance had been produced and this was the induced activity. Its atomic weight is less than that of the emanation, and at ordinary temperatures it is a solid. Since it disintegrates more rapidly than the emanation, it disappears again before any great quantity has collected. It is for this reason that the emanation or any other of the products of disintegrated radium are never found except in the presence of other radioactive materials.

The law of the slow disappearance of the radiation is therefore to be interpreted as follows: if, for example, a definite number of atoms of radium emanation N_0 are now present, then after the lapse of 3·83 days half of these have become disintegrated and, in place of them, the same number of atoms of the next substance, which is called Radium-A, have been produced. It is easy to discover the formula which will represent such a process. Let N_0 be the number of atoms of a radioactive substance when $t = 0$, and t seconds later let only N_t of them exist; then N_t/N_0 is a fraction representing the relative number now present. Since, according to observation, it has been found that after the lapse of equal times this fraction represents the number of atoms still present in terms of those which existed at the beginning of the particular time contemplated, we may write

$$\frac{N_t}{N_0} = e^{-\lambda t}, \text{ or } N_t = N_0 e^{-\lambda t} \quad . \quad . \quad . \quad (155)$$

where e is the base of natural logarithms.

From this it follows that the decrease in N in a small element of time dt is $dN = N_0 \lambda e^{-\lambda t}\, dt = \lambda N_t dt$. The decrease per second is therefore

$$\frac{dN}{dt} = N\lambda, \text{ or } \frac{1}{N} \cdot \frac{dN}{dt} = \lambda \quad . \quad . \quad . \quad (156)$$

The quantity λ is therefore a fraction representing the relative number of atoms which disintegrate every second. It can be calculated from a knowledge of the time in which the activity reaches its half-value period.

The definite time, T, at which this particular value is attained, *i.e.* when $N_t/N_0 = e^{-\lambda t}$ is one half, is given by

$$e^{\lambda T} = 2, \text{ or } T = \frac{\log_e 2}{\lambda} = \frac{0{\cdot}893}{\lambda} \quad . \quad . \quad . \quad (157)$$

This is the half-value period which is a characteristic of the various radioactive substances; *e.g.* it is 3·83 days for radium emanation.

Equation (155) really expresses the regularities which have been observed for whatever value we assign to t; if we increase it by the same amount T, the value of N will always be decreased by the same relative amount.

If the initial number N_0 is reduced to N_t after the lapse of t seconds, it follows that $N_0 - N_t$ is the number of atoms which have disintegrated in the same time; instead of this we may write $N_0(1 - e^{-\lambda t})$, which is thus the number of new atoms which have been formed.

If the product of the decay is radioactive, the process of disintegration will start again, as soon as a few atoms, *e.g.* those of radium emanation, have been formed. But since the number of disintegrating particles depends upon the number of atoms actually present, it follows that, at the beginning, only a very few of the emanation atoms will disintegrate. The formation of the emanation, however, depends upon the amount of radium present, and since radium only disintegrates very slowly—$T = 1580$ years for radium—the quantity which decays in a given number of days is constant, and it follows that the same amount of emanation will be produced each day. The quantity of emanation which decays each day is initially zero, but then begins to increase slowly; the amount continues to increase so long as the amount of emanation being formed is in excess of that which disappears. Ultimately a state of affairs is reached in which the quantity which disintegrates per second is exactly equal to that which has been produced in the same time. From now onwards, the quantity of emanation which is found with a definite amount of radium no longer changes. The emanation is said to be in radioactive equilibrium with the radium.

This amount of newly formed material depends upon the magnitude of λ, the disintegration constant, which varies from one substance to another. Let us suppose that we have a series of radioactive substances with constants $\lambda_1, \lambda_2, \lambda_3, \lambda_4 \ldots$; let N_1 be the number of atoms of the first substance present. We shall further assume that this substance disintegrates so slowly, that in the definite time in which equilibrium is attained, it has not altered appreciably in amount. Then in each unit of time a quantity $\lambda_1 N_1$ disintegrates. If there are N_2 atoms of the product of this decay, with a constant λ_2, then $\lambda_2 N_2$ is the number of its atoms which disintegrate per second. If this number is equal to $\lambda_1 N_1$, then just as many atoms of this substance will be formed as disintegrate per second, and the substance will be in equilibrium with the initial material. Then we have $\lambda_1 N_1 = \lambda_2 N_2$, or $\lambda_1 : \lambda_2 = N_2 : N_1$, and this ratio will remain unchanged. Even if the new product is radioactive, and likewise its product, the above considerations still apply. Since the disintegration constant is, according to equation (157), inversely proportional to the half-value period, it follows that the amounts of the various substances in equilibrium with one another are proportional to their half-value periods, *i.e.*

$$N_1 : N_2 : N_3 \ldots = T_1 : T_2 : T_3 \quad . \quad . \quad . \quad (158)$$

The proofs of the validity of this bold and revolutionary hypothesis are extraordinarily numerous and convincing. The production and decay of radioactive substances can be followed from day to day if the half-value period is numerically suitable, so that these facts can be confirmed.

From all radium preparations a gas is emitted continuously—it is the emanation, which can be isolated from the parent substance and examined. Year after year, after the parent substance has been left for the same number of days, the same amount of emanation is always produced. It can be shown that this emanation is a gas. **Ramsay** (1911) determined its density. This gas boils at $-63°$ *C.* and freezes at $-71°$ *C.* Its spectrum has been investigated and found to be different from that of all other bodies.

Alpha Particles. We have now arrived at the stage when it becomes necessary for us to discuss the nature of alpha particles. By measuring their deflection in electric and magnetic fields, and using equations (141) and (142), their velocity has been shown to be between $\frac{1}{20}$ and $\frac{1}{10}$ that of light, whilst the value of e/m for these rays is one-half that of a hydrogen ion in electrolysis. From these experiments another conclusion could be drawn—it concerns the mass of the particle. At first this conclusion was only probable, but it was soon established beyond all doubt. There were various grounds for supposing that the charge on an alpha particle was exactly twice that carried by a *H*-ion. If so, then the mass of an alpha particle must be four times as big as that of a hydrogen atom. The alpha particles were all equally large, and even when the alpha particles which were examined originated from different radioactive materials, they were still all equal to one another and had an atomic weight 4.

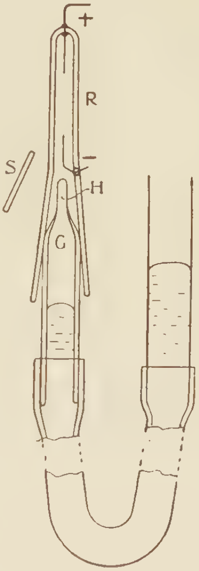


FIG. 146.

'Four' was also the atomic weight of one of the inert gases discovered by Ramsay (1895) in the air, and called helium. In 1903 he showed that the emanation contained helium. Soon afterwards it was discovered that a certain quantity of helium was occluded in all radioactive materials, and that this gas could be set free by heating the finely powdered material. All this pointed to the fact that there might be an intimate connexion between alpha particles and helium. In 1909 **Rutherford** and **Royds** showed in a very convincing manner that an alpha particle, after it had lost its positive charge, became an atom of helium. Their experiment was made possible because, in consequence of their enormous velocity, alpha particles were able to penetrate thin pieces of material which were impervious to ordinary gases. Rutherford and Royds had a thin glass cap *H*, Fig. 146, fused to the end of a tube *G* ; this cap was so thin that alpha particles could penetrate its walls. When he introduced radium emanation into the tube *G* and compressed it into *H* by means of mercury he was able to show that the alpha particles did actually pass through the glass by using the fluorescent screen *S* (zinc sulphide). The scintillations which appeared on this screen furnished the necessary proof (the tube *R* was removed during this part of the experiment). But helium gas was not able to pass through the walls of this cap, a point which he proved by filling *G* with helium instead of with the emanation. The presence of any helium in *R* could be detected by means of an electric discharge, but none was found on this latter occasion.

TABLE 29

The Radioactive Elements

Uranium-Radium Family.					Actinium Family.					Thorium Family.				
At.-wt.	Half-value period.	Rays.	Velocity.	Radioactive Substance.	At.-Wt.	Half-value period.	Rays.	Velocity.	Radioactive substance.	At.-Wt.	Half-value period.	Rays.	Velocity.	Radioactive substance.
238	$4\cdot5\times10^{9}$ yrs.	α	4·7	Uranium I										
234	24 days	βγ	48–60	Uranium X_1										
234	6·7 hrs.	β	—	Uranium Z										
234	1·14 min.	βγ	82–96	Uranium X_2										
234	ca. 10^{6} yrs.	α	4·9	UraniumII 3%										
230	76,000 yrs.	α	4·9	Ionium	230	24·6 hrs.	β	—	Uranium Y.	232	$16\cdot5\times10^{9}$yrs.	α	4·8	Thorium
226	1580 yrs.	α	5·0	Radium	230	20×10^{3} yrs.	α	5·2	Protactinium	228	6·7 yrs.	β	—	Mesothorium I
222	3·83 days	α	5·3	Ra-emanation	226	13·4 yrs.	β	—	Actinium	228	6·2 hrs.	βγ	37–99·8	Mesothorium II
218	3·05 min.	α	5·6	Radium A	226	18·9 yrs.	α βγ	5·6 38–65	Radio-actin.	228	1·9 yrs.	α β	5·3 47–51	Radio-thor.
214	26·8 min.	βγ	36–80	Radium B	222	11·2 days	α	5·5	Actinium X	224	3·64 days	α	5·5	Thorium X
214	19·5 min.	βγ α	44–99 5·2	Radium C 0·04%	218	3·92 secs.	α	6·0	Actin. Eman.	220	54·5 secs.	α	5·8	Thor.-Eman.
214	ca. 10^{-6} sec.	α	6·4	Ra.-C′	214	0·002 sec.	α	6·3	Actinium A.	216	0·14 sec.	α	6·0	Thorium A
210	1·32 min.	βγ	—	Ra-C″	210	36 min.	βγ	5·9	Actinium B	212	10·6 hrs.	βγ	63–77	Thorium B
210	25 yrs.	βγ	30–41	Radium D	210	2·16 min.	α β	6·0	Actinium C 99·8%	212	60·8 mins.	α β	5·7 99·9	Thorium C
210	4·85 days	βγ	77	Radium E	210	ca. 0·005 sec.	α	6·3	Act. C′	212	ca. 10^{-11} sec.	α	6·7	Thorium C′
210	136·5 days	α	5·3	Ra–F (polonium)	206	4·76 min.	βγ	60–92	Act. C″	208	3·2 min.	βγ	29–84	Thor. C″
206	∞	—	—	Ra–G (lead)	206	—	—	—	Act.–D (lead)	208	—	—	—	Thor. D (lead)

The velocity of the α- and β-rays is expressed as a percentage of that of light.

The velocity of the γ-rays is always equal to that of light.

He then introduced the emanation again into G, compressed it into H, having first ascertained that there was no helium present in R by showing that no discharge would take place through it. After the lapse of two days the presence of the first bright yellow line of the helium spectrum was observed, whilst after six days the whole helium spectrum was present. In this way it was established with certainty that the helium arose from the alpha particles which had passed through the walls of H.

For our present purpose it suffices to enumerate all the known radioactive substances in a table, which also contains the more important radioactive properties which these materials display [*cf.* Table 29, opposite].

Three series of active substances are known. Commencing with an initial material of high atomic weight, in each case a whole series of descendants are formed by disintegration, so that we can construct a kind of genealogical tree and speak of families of radioactive elements. The original parent of the longest known family is uranium, the last element of the periodic system with an atomic weight 238.

The *Ra*-family originates from uranium. Another family begins with thorium, the second heaviest of the elements in the periodic system, of atomic weight 232. The third family is a branch from the uranium family, inasmuch as II breaks up partly into ionium and partly into UY, from which latter substance the actinium series is derived.

4. DEDUCTIONS FROM RADIOACTIVE PHENOMENA.

A resumé will now be given of the results which have been derived from the study of this relatively new branch of physical research. In part, they are results which have confirmed our earlier ideas, while some have completed and extended them.

Modern Confirmations of the Atomic Theory. If the atomic theory is true, then all chemical changes must be the resultant of individual processes taking place between individual atoms or molecules. Hitherto it was only possible to infer this from certain indications, since all actions taking place between the individual atoms are so small in comparison with any one of our senses, that they remain imperceptible to us. The disintegration of radioactive bodies, however, is accompanied by such large exchanges of energy that one sees the possibility of making such actions perceptible to us. During recent years this feat has been accomplished.

(*a*) **Counting the α-particles.** It has been found that an ejected α-particle can, according to the velocity it possesses, produce 125–273 thousands of pairs of ions in air. If the charge on all these ions is used to discharge an electroscope, then providing the instrument is sufficiently sensitive, this charge can be detected. **Rutherford,** who first made these experiments, was not successful in these endeavours until 1908. He brought these thousands of ions into a strong electric field, in which they acquired such an enormous acceleration that they were able to ionize the air through which they subsequently passed. If each ion produces, let us say, a thousand new ones, then the total charge carried by the ions of one sort is

$$4{\cdot}77 \times 10^{-10} \times 1000 \times 200{,}000 \doteqdot 0{\cdot}1\ E.S.U.$$

When this charge is given to a sphere whose radius is 1 cm., its potential becomes 30 volts. This can be detected conveniently with the aid of a moderately sensitive electrometer. If the arrival of a single α-particle is to be detected, however, the electrometer needle must return to rest as soon as possible after it has been deflected. It is for this reason that string electrometers are generally used to-day, since the above-mentioned return occurs in less than 0·1 sec.

These phenomena can be demonstrated very easily by means of a device due to **Geiger.** K is the essential part of the instrument : it consists of a small metallic chamber in which there is a sharp point, Fig. 147. It is called an ionization chamber. This inside point, which may be a gramophone needle, is connected to the inside coating of a Leyden jar raised to a potential of 3000 volts. Each time an α-particle enters the opening of this vessel a small charge will be released from the point of the needle. All the ions of one sign, their number being increased by collision with molecules of the gas, pass to the walls of K, to which they give up their electric charges. This discharge can be shown very easily by inserting a sensitive telephone between this chamber and the external coating of the Leyden jar. If, instead of a telephone, a quick-moving electrometer E is used, then the arrival of each individual particle can be registered on a moving piece of photographic paper, in much the same way as a telephone message may be received. Fig. 148 shows a portion of such a record taken with the aid of a 'universal electroscope' [*cf.* p. 234]. The distance between two vertical lines indicates a period of one second. These experiments show that the radiation from radioactive substances does not take place continuously, but that the particles are emitted one after another. These particles are helium atoms, so that it is possible to demonstrate the atomic structure of matter to an audience.

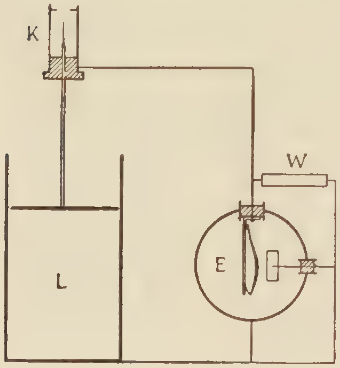


Fig. 147.

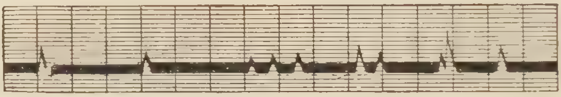


Fig. 148.

But these experiments are still more important in the realm of pure science. When this process of disintegration is recorded photographically, it becomes possible for a count to be made of the number of particles emitted by radium in a given time. In combination with other facts, this leads to much new information concerning the atom.

First, let us consider the method of counting the number of α-particles ejected by 1 grm. of radium per second. All the α-particles will most certainly not be allowed to enter the chamber. If the distance of the opening from the radium is R cm., Fig. 149, whilst r is the radius of the hole in the chamber, then the area of the sphere R is $4\pi R^2$, whilst that of the opening is πr^2 : the ratio of the area of the small

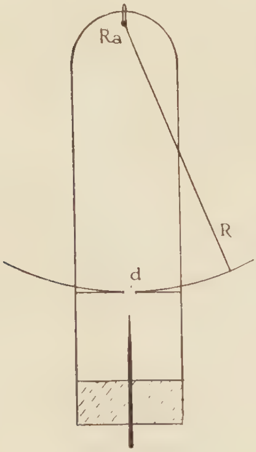


Fig. 149.

opening to that of the whole sphere is therefore $r^2/4R^2$. Since the α-particles are ejected in all directions from the radioactive preparation, it follows that this fraction represents the relative number of particles passing into the detecting device. It is therefore only necessary to multiply the number of particles passing into the chamber by $\frac{4R^2}{r^2}$ in order to obtain the total number of particles ejected by the radium.

If, moreover, by weighing and chemical tests of its purity, it is known how much radium is actually present, then the number of α-particles sent out from 1 grm. of radium per second is also known. According to the measurements of **Hess** and **Lawson** this number is $3{\cdot}7 \times 10^{10}$.

When, in addition to this, the value of the decay constant λ for radium is known from electrometer measurements, then the proportion of the whole which disintegrates every second is also known. The number of atoms in that fraction of 1 grm. which has decayed in one second is $3{\cdot}7 \times 10^{10}$. The number of atoms in a gram-atom or 226 grm. of radium can therefore be found at once. If the radioactive constant is λ, and N is the number of atoms in 226 grm., then

$$N\lambda = 226 \times 3{\cdot}7 \times 10^{10}.$$

Since λ has been found to be $1{\cdot}4 \times 10^{-11}$, we have

$$N = \frac{226 \times 3{\cdot}7 \times 10^{10}}{1{\cdot}4 \times 10^{-11}} = 60 \times 10^{22},$$

as we have previously found.

Even without knowing the radioactive constant, this number can be ascertained as follows. Since each α-particle becomes an atom of helium when it has lost its charge, it follows that a gram of radium produces $3{\cdot}7 \times 10^{10}$ helium atoms per second. Since radium disintegrates so slowly, we may assume that this number is constant over a period of one year. A year consists of $31{\cdot}5 \times 10^6$ seconds, so that the number of atoms produced in this time is

$$3{\cdot}7 \times 10^{10} \times 31{\cdot}5 \times 10^6 = 1{\cdot}17 \times 10^{18}.$$

The helium originating from a strong radium preparation in one year was then collected and measured ; it amounted to $0{\cdot}042$ cm.^{2} per grm. This quantity of gas contains the above-mentioned number of atoms. Consequently the number of atoms in $22{\cdot}4$ litres, which is the volume of one gram molecule, is

$$N = 1{\cdot}17 \times 10^{18} \times 22{,}400/0{\cdot}042 = 62 \times 10^{22}.$$

Considering the difficulties which attend the measurement of such small quantities of gas, the consistency of these two results is very satisfactory.

(*b*) **The Tracks of α-particles.** When a space saturated with water vapour is suddenly cooled, a portion of the vapour separates out in the form of liquid water. It has been found that this separation always occurs in the form of fine droplets, each of which is found round some nucleus. This nucleus is usually provided by the dust particles which are ever present in large numbers in the air. (This is why the rain cleanses the air and makes the windows dirty.) But if the air is freed from dust, it has been shown that the ions present in the air can act as nuclei for this condensation. Now if it is true that individual rays are sent out with a great speed from a radioactive substance, and that these ionize the air along the path they traverse, it follows that these ions must be situated on the path which the particle has taken ; if this path can be rendered visible, then the whole track of one α-particle can be observed. These tracks can be made visible by utilizing the property in virtue of which ions are able to act as nuclei around which a vapour may condense. For if the air is suddenly cooled by allowing it to expand a little, then condensation takes place along the path which the ions occupy; when illuminated by a bright light, this path is easily seen. The path, however, will only appear sharp for a moment, for afterwards the ions begin to diffuse. **C. T. R. Wilson,** who first performed this experiment, succeeded in obtaining a photographic record by causing a bright light to illuminate the vapour at the very instant that the expansion took place. Not only were the individual tracks thus seen, but it was also observed that single particles collided with the air molecules. This was especially so near the end of the range of an α-particle where it was travelling more slowly—when such an encounter occurs the track often shows a distinct bend [*cf.* Fig. 150].

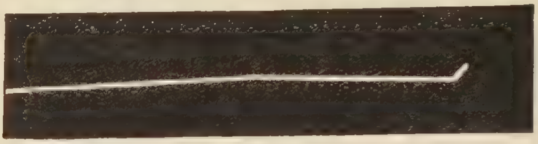


Fig. 150.

By means of the simple apparatus shown in Fig. 151 the above experiment is very easily repeated. First by means of an air pump (a water-pump will do), the air in the flask *F* is partly removed. When the wide stop-cock *H* is suddenly opened some of the air from the chamber *K* rushes into this flask, thereby cooling itself so that condensation can take place. Inside the chamber is a needle covered with a trace of radium. The tracks of the particles which are sent out at this instant appear bright in front of the dark background, if the condensation chamber is brilliantly illuminated either from above or from the side. It may therefore be said that the emission of an α-particle from a radioactive substance can be classified amongst facts which have been observed.

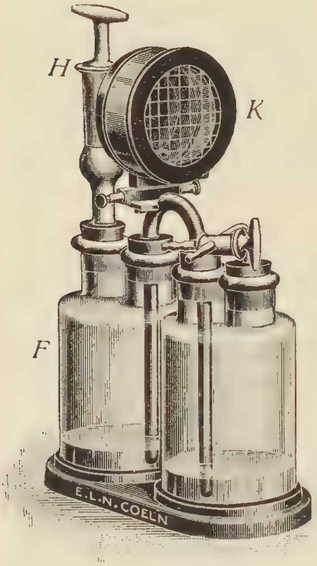


Fig. 151.

Recent Knowledge from Radioactive Phenomena. We come now to a few results obtained from a study of radioactive bodies, which have not only fortified our existing knowledge, but which have also extended it by allowing us to obtain a deeper insight into the structure of matter.

In the first place, a whole series of new substances appear. An attempt was made to classify them according to the periodic system. Such an attempt was necessary, for if the system does not include all elements, without exception, it is no longer a system of classification. Besides, uranium and thorium had already been assigned to places in this system before they were shown to be radioactive. Likewise helium, which is always produced during a radioactive transformation, had its own position along with the other inert gases in the periodic table. The task of inserting the radioactive elements in this table was rendered easier by the fact, that, since the atomic weight of the alpha particles was known, the probable atomic weight of a product could be inferred from the order in which this new substance was formed from the parent element.

One difficulty which arose was due to the fact that when a beta particle was sent out, the new substance had an atomic weight which was practically equal to that of the parent element. Further, since it had been shown that the actinium series was an offshoot from the uranium series, it was clear that several elements from one or other family might have the same atomic weight, as a glance at Table 29 (p. 316) will show. Moreover, no means were found whereby these several substances, which must be different from one another, since their disintegration occurred at different rates, could be separated ; this meant that these substances had identically the same chemical properties. If radioactive phenomena had not indicated quite clearly that there were two distinct substances present, every chemist would have said that only one substance was being dealt with. Here, for the first time, elements were found which, in spite of their different atomic weights, had the same chemical properties. These substances had already been called isotopes [*cf.* p. 217]. We have already seen the fruits of this hypothesis. The discovery of isotopes was really due to researches made on radioactive changes, and we have seen how single places in the periodic system of the elements [*cf.* Table 5, p. 111] may be filled by several bodies with different atomic weights.

Hitherto the elements had been regarded as something which never altered. The law of the conservation of matter had been held to apply to mass in bulk and even to the various atoms themselves. Indeed, considering the fact that every compound can be decomposed into the same number and types of atoms as first united together in order to form it, we had to assume that the elements were indestructible, for there were no sound reasons why a contrary view should be held. Now we see that all this is not quite so. For a series of elements, observations have been made which show that they are constantly and regularly disappearing. Others are immediately formed from them. The heaviest elements in our system, uranium and thorium, are always disintegrating. These two elements decay very slowly, but with others the process can be watched. The amount of the decay can be measured, as well as the amount of helium which is produced.

It is evident that these new facts will cause us to change our conception of the whole material world.—Of the whole ? And yet we only have proofs applying to a few elements.—It is true that our evidence is furnished by a few elements, but we will here remind ourselves of what has been said concerning reasons which have been proved, and those which are merely convincing. Is the creation of one element from another only to be a characteristic of some few elements ? It is difficult to imagine such a state of affairs. Certainly the various elements have different properties, but are their fundamental properties to be so different that some elements originate from others and form one family, whilst others do not ? And yet are all these different elements to form one comprehensive system with one another ? And in this same system are all elements with essentially different origins to stand side by side with one another ?

Besides, it is not true that, in dealing with this question, we have no actual clues upon which our assertions can be based. All radioactive elements disintegrate, and, without exception, in so doing they emit either a beta particle which has practically no mass, or else a helium atom whose atomic weight is four. Hence it follows that the atomic weights of all elements which are formed from one another after an alpha particle has been ejected must differ by four units. If therefore the remaining elements have been formed as a result of atomic disintegration, then the differences in their atomic weights must either be four, or an integral multiple of four. When we look at the table of atomic weights on page 108, we find that this occurs many times. Not only is this approximately so, but most of the elements have atomic weights which are integral—since the discovery of isotopes we know that the atomic weights are really the mean values of isotopic weights, which are always integral. The numbers 4, 12, 16, 20, 24, 28, 32, 36, 40, strike us at once when we look at the table. The series 19, 23, 27, 31, 35, is also to be observed. It would be very peculiar if this interval 4 should occur repeatedly throughout the periodic table, in many instances an alpha particle being responsible, whilst in others it should happen quite by chance. Supported by this fact that differences in atomic weight are often constant, we ought certainly to regard as one which is very probable the assumption that all the elements have been formed in the same way from elements of high atomic weight. If so, then our whole system certainly becomes still more important. All the known elements are characterized by a certain uniformity and interrelationship. The elements form one big family ; they originate from one another. Even if the details concerning this interrelationship cannot always be discerned, or perhaps one should say understood, we no longer wonder at the fact that neighbouring elements should possess similar properties, *i.e.* in short, that all the elements form one large system.

From this hypothesis we should be able to find the reason why some elements occur more frequently than others. It is only those radioactive substances which have a long existence that can be obtained in any appreciable amount. The greater frequency with which other elements occur must therefore be due to the fact that they have a much longer life. This frequency is a function of the rate of decay, as expressed by equation (158).

If this hypothesis is true, it follows that the ultimate destiny of all matter will be quite similar to that of energy as deduced from the second law of thermodynamics. We have seen that all elements are undergoing a process of decay, and that from the heavier elements lighter elements are formed : this leads us to make the following inference. If at some place the process is not occurring in the reverse direction, then in the far-distant future all matter will have decayed. The elements of higher atomic weight are gradually diminishing, whilst only the lighter ones remain, and we have no indication at all that the process is taking place in an opposite direction.

Radioactive phenomena permit us to make yet another step into the innermost parts of the material world. Hitherto we have only spoken of the atom as the ultimate and smallest of material particles. Here, however, we discern phenomena which show us that the atom ought not to be regarded as an entity having no structure, as we generally regard a lead shot or drop of water. In the first place we see that an enormous amount of energy resides inside radioactive atoms. This energy appears during the process of disintegration, being converted into other forms of energy, and in most instances passing at once into thermal energy. The velocity of most alpha particles is about one-twentieth that of light, *i.e.* $1{\cdot}5 \times 10^9$ cm. per sec., or 15 million metres per second. If a shot has imparted to it an initial velocity of 1000 metres per second, then that of an alpha particle is 15,000 times greater. The most noticeable result, however, is the following. If, as a result of an explosion, say of an atom of radium, an atom of emanation has been formed, then this new atom can also explode. After not quite four days half of these atoms have exploded, again hurling forth exactly the same projectile, viz. a helium atom. The product is called radium *A*, and after three minutes half the radium *A* atoms have exploded again, hurling forth a helium atom with a similar enormously high velocity. If this energy could be used as required, *i.e.* if we could control the process of disintegration at will, we might think that the energy of radioactive bodies could be used for various useful purposes, for it is very great. According to the measurements of different workers, 1 grm. of uranium in equilibrium with all its products develops 10^{-4} grm.-cals. per hour. Since the amount of uranium which disintegrates in this time is only $17{\cdot}6 \times 10^{-15}$ of the whole, it follows that each gram of uranium, in decaying, emits $5{\cdot}7 \times 10^9$ cals. Now 1 grm. of coal on the average supplies 8000 cals., so that 700,000 grm. of coal would have to be burnt in order to supply the same amount of heat as 1 grm. of uranium. But a bird in the hand is worth two in the bush! We are not able to obtain this energy. We are in a somewhat similar position to the owner of much money who, not being allowed to spend more than an infinitesimal amount per annum, must needs perish with hunger.

Amongst all the radioactive bodies there are 23 which send out alpha particles. Since all these α-particles have the same atomic weight 4, we have to conclude that such particles really exist as separate entities inside the parent substance, *i.e.* the mass of each helium atom is held together more closely than it is to the remaining portion of any atom which emits an alpha particle. This is really a very obvious assumption, for this entity must certainly be formed inside the atom before it can be ejected. When we remind ourselves of the fact that the difference in the atomic weights of neighbouring atoms is frequently equal to the atomic weight of helium, viz. 4, we have to conclude that, in general, a helium atom is an element of construction from which the heavier atoms are built, whilst the helium atom itself may consist of four hydrogen atoms.

This hypothesis has been confirmed again during recent years. **Rutherford** imagined that, if the atoms were constituted of smaller parts, it should be possible to disintegrate them, if one could get enormously large forces to act upon them. For this purpose it is not sufficient to allow such a force to act upon a large extensive mass, it must exert its influence upon an individual atom. The greatest such force with which we are acquainted is that which an alpha particle can exert in virtue of its high speed. Rutherford therefore allowed alpha particles to pass into a space containing nitrogen. When this happens, a head-on collision will sometimes take place between an alpha particle and a nitrogen atom. Will the nitrogen atom be able to withstand this impact? Rutherford showed that under the influence of such collisions new rays appeared; these had such a range that it was necessary to assign to them an atomic weight equal to unity, *i.e.* they were atoms of hydrogen. These particles were detected by the occasional flash which they produced on a zinc sulphide screen placed at a distance beyond the range of the alpha particles. Rutherford, and other workers who were able to confirm these findings, have allowed alpha particles to exert their influence on other elements, and for a whole series of them it has been found that under this action hydrogen is produced. Amongst 10^6 ejected alpha particles there is only one which hits another atom in a head-on collision, *i.e.* only one atom of hydrogen is produced. The range of this hydrogen atom is several decimetres in air at ordinary pressures. Its energy is greater than that of the colliding alpha particle, so that it is necessary to conclude that this energy has been acquired from the store of energy inside the nitrogen atom.

Using a strong thorium *C* preparation, **Meitner** has been able to photograph alpha particles after the manner used by **C. T. R. Wilson,** and has succeeded in detecting the path of a hydrogen atom which had been formed under these conditions (Fig. 152). **Pettersson** and **Kirsch,** working at the Radium Institute in Vienna, have extended these experiments to all possible elements, and so far have been able to show that an atom of hydrogen is ejected from 27 of the elements investigated, including carbon and oxygen. For the sake of brevity we shall simply denote the elements by their atomic numbers. The elements which have been caused to disintegrate artificially are all those from 3 to 19, and then 22, 24, 26, 29, 34, 35, 40, 50, 52 and 53 [*cf.* Table 4, p. 108].

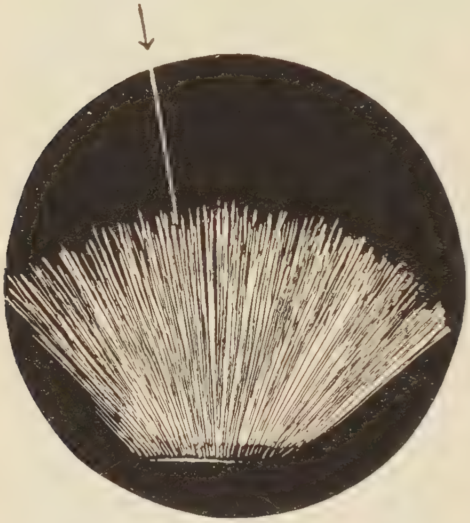


Fig. 152.

Since helium atoms always manage to survive the effects of radioactive explosions, we are forced to assign to them a very high stability, *i.e.* their constituents must be held very firmly together. On the other hand, it is seen that even atoms of elements whose atomic weights are 12 and 16 yield hydrogen when an alpha particle makes a head-on collision with them. From this we conclude with some probability that not all the hydrogen atoms are united in helium atoms, but that some are present individually in atoms of carbon and oxygen.

Although the alchemists tried in vain to transmute one metal into another and to find how gold could be made, yet in this phenomenon of the disintegration of atoms we do actually see the tendency for one element to pass into another, but this tendency is always in one direction ; atoms only break up, they are never built up by this process.

Rutherford's Atom. In 1911 **Rutherford** attempted to form a picture of an atom from the facts which we have already discussed, and from a whole series of others. Rutherford's idea of an atom has since become very important ; other workers have enlarged upon it, whilst new facts have been discovered which only serve to confirm his view. It is therefore necessary for us to discuss this theory in brief. In 1909 **Geiger** and **Marsden** made some experiments on the passage of α-rays through thin metal sheets. The results which they obtained were very similar to those which **Lenard** had already found in connexion with the passage of cathode rays through metal foil [*cf.* p. 306]. In general, the α-particles are deviated a little from their straight path when they pass through aluminium or gold leaf; in consequence, the whole beam tends to spread slightly outwards. A few isolated particles, however, are deflected to such a large extent that they appear on that side of the leaf upon which they had originally fallen [*cf.* Fig. 153]. Out of 8000 particles only one rebounds in this manner. It is impossible to think that all the particles experience the same resistance whilst in the metal, *i.e.* the metal is not uniformly constituted throughout. On the contrary, it is more probable that at certain points in the metal some great force may be present. These points do not occur very frequently in comparison with others, where no such forces arise. At these few points, however, very large masses must reside, since they are able to deflect the rapidly moving α-particle to one side. Actually the projectiles which collide with these points rebound like indiarubber balls from a wall.

Moreover, in atoms, we have both positively and negatively charged particles. Radioactive processes have shown that the positive charge is found only in connexion with α-rays whose mass is equal to that of four hydrogen atoms. The negative charge is on the electrons whose mass is vanishingly small in comparison with that of the atom. Rutherford therefore assumes that an atom consists of two parts—the relatively heavy portion which carries the positive charge, and the negatively charged portion which is practically without mass. These charges are always equal since, as a whole, the atom appears to be uncharged. But if the inside of an atom is to be the seat of enormous forces, then these two opposite types of charge cannot actually be combined, for otherwise their effects would be annihilated. Hence, in spite of the tendency which opposite charges possess to coalesce, something must be present which prevents them from so doing. Rutherford assumed that the light electrons encircled the heavy positive nucleus. This nucleus contains the whole gravitational mass of the atoms and is built up of helium and hydrogen atoms. It also contains the whole of the positive charge, and in spite of its mass, it only occupies a very tiny region in space. This is why the α-particles are able to approach it very closely and, in so doing, are repelled by very great forces according to Coulomb's law [*cf.* p. 234]. This, however, takes place very rarely, since the nucleus is exceedingly small. Around the nucleus the electrons move in circles at various distances from it; the negative charge on an atom is therefore distributed around the nucleus in the form of a cloud of electrons. The sphere of action of an atom embodies the region in which these electrons move; the diameter of this sphere will be the diameter of the atom. From the small probability that an alpha particle may be expected to rebound, as also from the small number of nitrogen atoms which are destroyed by a head-on collision, and from the enormous force which must be present in order to cause a rapidly moving α-particle to rebound, it can be estimated that an α-particle is permitted to approach the nucleus of an atom to within about 10^{-12} cm. From reasons

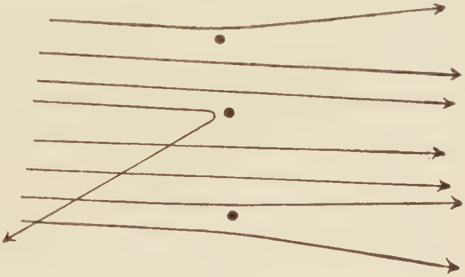


Fig. 153.

based on this and similar experiments, it may be assumed that the radius of a nucleus varies from 3×10^{-13} cm. for helium to 6×10^{-12} cm. for radium. According to this hypothesis, an atom, consisting of a central nucleus encircled by electrons, is not unlike the solar system, which consists of the sun and its attendant satellites (the planets).

As observations on single disintegrating atoms show, the laws of disintegration are purely statistical ones, which can only apply to large numbers and large intervals of time. The individual phenomena, which in this case we are also able to observe, do not support these laws, but manifest themselves apparently irregularly. Our whole conception of Nature would be contradicted if these individual processes were not governed by some law, but at present we have no evidence as to why the atoms break up as they do. 'If a radium atom disintegrates again after it has been formed from ionium, it may do so within several minutes ; another may live for several days, whilst another may survive for millions of years before it passes into gaseous emanation ; after this a whole series of products are formed in rapid succession, until finally the process ceases, when an atom of lead has been formed. What is the cause of this differing behaviour ? We do not know.' (**Planck.**) And thus this illuminating chapter in Physics terminates with a new problem which is not to be solved until we have much more exact knowledge than we possess at present concerning the processes at work inside and the structure of the nucleus.

CHAPTER VIII

ELECTROMAGNETIC WAVES

RESUMÉ. The discharge of electricity, let us say from a Leyden jar, occurs under certain conditions in the form of oscillations. In the region round about, these oscillations are able to produce electric waves which are propagated over great distances, similar to the vibrations of a tuning fork which excite the air so that we hear a note, or like the light which comes to us from remote stars and stimulates our optic nerve. If such electric waves are produced by the discharge of Leyden jars, or other bodies of similar dimensions, then their wave-length is very large, extending in some cases to several metres. But it has also been demonstrated that the intense radiation emitted by hot bodies is also produced by the electrical oscillations of small electric charges. Finally, it has been shown that Röntgen rays and the γ-rays from radioactive bodies consist of electric waves which only differ from one another by their wave-length, just as a high note differs from a low one. In this way electric waves cover a range which is very much greater than that of sound waves. The visible rays, by means of which we are able to discern things directly, only occupy a small fraction of this range. These short waves, especially those which are emitted from the smallest particles of matter, have enabled us during the last decade to make many inferences with respect to the internal structure of atoms. We will first give a synopsis of all electrical rays and their wave-lengths: in order to compare them with waves of sound we shall also state the number of octaves which they would occupy if we use the term octave to denote rays having a wave-length one-half that of another, as in the case of sound waves.

TABLE 30

Synopsis of all known Electromagnetic Waves

Nature of the Radiation.	λ.	Octaves.	ν/sec.	$h\nu$ in ergs.
Wireless waves, longest . ca.	30 km.	28	10^4	$6{\cdot}5 \times 10^{-23}$
,, ,, shortest . ca.	3 m.		10^8	$6{\cdot}5 \times 10^{-19}$
,, ,, (laboratory) ca.	0·1 mm.		3×10^{12}	$19{\cdot}6 \times 10^{-15}$
Infra-red rays, longest . . .	0·3 mm.	8	10^{12}	$6{\cdot}5 \times 10^{-15}$
Visible ,, ,, red . .	0·8 μ	1	400×10^{12}	$2{\cdot}45 \times 10^{-12}$
,, ,, shortest, violet.	0·4 μ	8	800×10^{12}	$4{\cdot}9 \times 10^{-12}$
Ultra-violet rays, shortest .	1·22 $\mu\mu$		250×10^{15}	$16{\cdot}3 \times 10^{-10}$
Röntgen rays, longest . .	66 $\mu\mu$	17	$4{\cdot}5 \times 10^{15}$	$39{\cdot}4 \times 10^{-12}$
,, ,, shortest . .	0·0057 $\mu\mu$		50×10^{18}	$3{\cdot}5 \times 10^{-7}$
Gamma rays, longest (*Ra–Act*)	0·039 $\mu\mu$		8×10^{18}	$5{\cdot}0 \times 10^{-8}$
,, ,, shortest (*Ra–C*)	0·0005 $\mu\mu$		600×10^{18}	$3{\cdot}9 \times 10^{-6}$
Cosmic rays	0·000004 $\mu\mu$ (= 4×10^{-12} mm.)		750×10^{20}	$4{\cdot}9 \times 10^{-4}$

The frequency ν is connected with the wave-length λ by means of the equation $\nu\lambda = c = 3 \times 10^{10}$ cm. per sec. The last column contains the quantum of energy associated with each wave-length according to the quantum theory [*cf.* p. 337].

Even if it has been proved that all these waves are essentially alike, consisting as they do of electromagnetic waves, yet the methods whereby they may be produced, or the means whereby they may be detected, vary very much according to the wave-length. For this reason each type of wave will have to be discussed alone. But since it is only the excitation of these waves which differs according to the wave-length, and not, or at least not fundamentally, the manner in which they are propagated, we shall at first only speak of the production of these waves. The perception of these waves would certainly be impossible if they were not propagated through space, but the laws which govern their motion will be discussed collectively for all of them, in the last section of this book.

1. ELECTROMAGNETIC WAVES

We shall begin our discussion with those waves which are produced by electrical discharges, and whose electrical nature is at once apparent from the manner in which they are produced. In the first place their wave-length is very long. These rays, together with all their properties, were described in a very remarkable way by **Maxwell** (1865) before any physicist had ever detected such waves. As we have already remarked, Maxwell concluded from the fact that the ratio of the electromagnetic unit of current to the electrostatic is equal to the velocity of light (3×10^{10} cm. per sec.), that the transmission of light is an electrical process, and that electrical disturbances travel through space in the form of waves, and that these oscillations are identically like those of light. From the laws of electrostatic and electromagnetic induction he developed the whole theory of electric oscillations and their propagation in space. He was able to show that they would be refracted, reflected, polarized and propagated just like light waves.

In the year 1888 **H. Hertz** (1857–1894) endeavoured to produce electric waves and demonstrate their properties. We have already seen above [*cf.* p. 267] how the discharge of a Leyden jar may be made oscillatory by the combined action of capacity and self-inductance. But even with a single piece of wire which is bent and then charged so that a spark passes between its two ends, the discharge takes place in a similar way, only the frequency is very much increased. The oscillations only cease when all the electrical energy has been converted into heat according to Joule's law.

The period, T, of a mechanically produced oscillation has been shown to be expressed by $T = 2\pi\sqrt{1/a}$, where a is the acceleration at unit distance from the zero position. A very similar formula now applies, which we shall not prove but only deduce by analogy. In mechanics, the acceleration is defined by means of the force f and the inert mass m, by the equation $a = f/m$. In the present problem the force corresponds to the electric potential V, which, according to equation (108), is given by $V = q/C$, so that when the charge is unity we have $V = 1/C$. In place of the mass, we have the self-inductance L, so that istead of f/m we now have $\frac{1}{CL}$. When this value is inserted in equation (27) we obtain

$$T = 2\pi\sqrt{LC} \qquad . \quad . \quad . \quad . \quad . \quad (159)$$

The exact theory leads to this same formula which experience has only served to confirm completely.

The Excitation of Electric Waves. The fact that the discharge from a Leyden jar was oscillatory had been shown by **Feddersen** in 1859; by using a rotating mirror he was able to show that the discharge consisted of a series of sparks passing to and fro between the terminals. Hertz now proved that these oscillations produce waves which are transmitted through space. He showed that they were reflected by bodies upon which they were incident, that they were refracted by a prism of pitch, and that they were polarized just like waves of light. As a detector of these waves Hertz used a second piece of wire shaped like the first, so that it too possessed a spark-gap. When the first loop was excited sparks could be seen passing between the extremities of the second loop, in the dark, by a well-rested eye. This sparking was observed even when the two loops were several metres apart. Such a loop constituted the first apparatus for the reception of electrical waves. Since then exactly forty years have elapsed. To-day electrical waves encircle the whole earth day and night. They carry, not only decipherable signs, but also the spoken word, the sound of the human voice, across the oceans.

Hertz also caused these waves to interfere with one another, so that he was able to deduce their wave-length. From the known dimensions of his transmission apparatus he calculated their periodic time, so that by means of equation (43), $\lambda = cT$, he determined the velocity c with which these waves were propagated. Later, and more refined methods, have shown that $c = 300{,}000$ km. per sec., a value which is exactly equal to the velocity of light. If we remind ourselves of the things which have been said previously concerning the velocity of waves [*cf.* p. 80], we shall understand the importance of this agreement. Light waves and electric waves require the same medium for their propagation. Viewed from one side only, this fact supplies the evidence from which we deduce that electrical waves are light waves, for, like these, they can be reflected, refracted, and also made to exhibit the phenomena of interference and polarization.

At first this was only established for the transmission of waves in air, which, for optical purposes, is practically like a vacuum. But light can also be propagated through water, glass, and other media, and, as we have already deduced from the general theory of waves, the velocity with which it is transmitted varies with the nature of the medium. The behaviour of electric waves as they pass through prisms made of pitch, as was revealed by the experiments of Hertz, indicates that they behave in a similar way, a fact which Maxwell had already established theoretically. The following considerations will enable us to appreciate this subject more fully. According to equation (46) [*cf.* p. 80] the velocity with which a wave travels through a medium depends upon the elasticity and density of the medium; in fact, it is equal to the square root of the ratio of these quantities. That different media have different elasticities means that the force with which a medium returns to its zero position after being disturbed, varies from one medium to another. For electric disturbances, we have shown [*cf.* p. 242] that the restoring force is determined by the dielectric constant. We shall therefore not be surprised to find that, as the exact theory shows, the velocity of propagation of electrical waves is proportional to the square root of the dielectric constant. For light waves, the velocities with which they move through various media vary as the refractive index, according to equation (48) [*cf.* p. 86].

Accordingly the square root of the dielectric constant plays the same rôle in connexion with electromagnetic waves, as does the refractive index for light waves; the formal development of this theory has shown that these two quantities are numerically equal, a fact which later experiments have confirmed, as the following table indicates.

TABLE 31

	Substance.	$\sqrt{k}$.	n.
Gases	Air	1·000295	1·000294
	Hydrogen	1·000132	1·000138
	Carbon monoxide	1·000345	1·000340
	Carbon dioxide	1·000473	1·000449
Liquids or solids	Benzene	1·514	1·506
	Petroleum	1·385	1·386
	Calcite	2·90	2·7
	Alum	2·58	2·5
	Ebonite	1·59	1·6

The agreement which is thus found to exist for all the important properties is the evidence which shows that electric waves and light waves are identical.

The Effect of Magnetic and Electric Fields on Light Waves. From the other side of the problem equally weighty evidence is also obtained. If electromagnetic waves are light waves, then, as Faraday had already intimated, light waves must be electromagnetic. One should therefore be able to demonstrate the effect of electromagnetic fields on light waves. In order to appreciate fully the conclusiveness of the following, let us again remind ourselves of what has been said already concerning action and reaction. It is only possible for forces of the same nature to affect one another. If, therefore, it can be shown that light waves are influenced by electric and magnetic forces, then such forces must be associated with the emission of light. Now in 1896 **P. Zeeman**, working in Amsterdam, showed that the rays of light emitted from a sodium flame were changed when the flame was placed in a strong magnetic field. At first the lines were only broadened. Later and more refined experiments have shown that when the light travels perpendicularly to the lines of force each spectrum line is accompanied by two others, one on either side, corresponding to a shorter and longer period of oscillation respectively. When the light travels along the lines of force, then only two lines appear, one corresponding to a shorter and the other to a longer period than that of the original wave. At this time electrons had just been discovered and measured, so that **H. A. Lorentz** (1853–1928) applied to this problem the theory of electrons, which are those small particles of electricity from which all light waves must originate. Zeeman, supported in his work by the theory which Lorentz had developed, determined the ratio ε/m for these rays and found that it was exactly the same as for cathode rays. In this way it was not only shown that in dealing with light we are really concerned with electromagnetic oscillations, but also that the electron is the originator of such disturbances.

Some year later **Stark** attempted to detect the influence of electrostatic fields on light waves. His experiments were carried out at Aix-la-Chapelle, and in 1913 his attempts met with success. The difficulty in these experiments was to obtain a sufficiently strong electrostatic field to act on the electron just as it emits light. Stark succeeded eventually by producing a glow in a gas by permitting cathode rays to impinge upon its molecules. Accompanying the line in the original spectrum he found others, some having a longer, and others a shorter period.

The experiment closed the ring completely, for it had now been shown quite clearly that electromagnetic waves are light waves, and light waves are electromagnetic waves. The establishment of this fact marks a very important step forward in our knowledge of matter. If our aim is to refer many various phenomena to the least possible, but most comprehensive laws, then the correlating of these two apparently divergent branches of science is a great step in this direction. The knowledge that light oscillations are electrical processes has brought about a very rapid advance in science during more recent years: these developments have yet to be discussed in the next few sections.

But the reader should be careful not to imagine that everything which had been discovered before in connexion with rays must now be useless or declared inexact, for, once a fact has been established, it can never be refuted. We can understand that an explanation or hypothesis, given at some time in connexion with some fact, may later be shown to be untenable, but the improvement will only be made on our erring conception, and not on the facts. The laws of light and heat waves, which had been discovered a long time previously, are still valuable to-day, and many of them, in conjunction with others, have led us, even in these days, to make other important conclusions.

2. Heat Rays

The Relationship between Heat and Light Rays. Even when one is standing at some distance away from a furnace one can feel the heat which is emitted; experience has taught us that a piece of paper will screen us from this heat. For many years it had been believed that a kind of heat-substance, caloric, was given out by the furnace. It was only at a later date, when the electric, magnetic, and other fluids had lost their supporters, and when heat had become recognized as a state of energy, that the conviction arose that the energy which is received from a hot body comes to us in the form of waves.

But even then the heat rays were regarded as a special type of ray, accompanying the light and chemical (actinic) rays emitted from the sun. This theory was supported apparently by the fact that heat rays sent out by a hot body are not necessarily accompanied by light rays, that they produce in us a sensation of warmth, and that they are capable of warming a thermometer in the same way as they warm our hands.

Apart from the circumstance, however, that they are emitted by bodies at high temperatures, heat rays have no more and no less to do with heat, than the waves which carry a message across the ocean, which are responsible for the colours of the evening sky, or which affect a photographic plate. In particular, it was found that rays of any wave-length, such as we have indicated in Table 30 [*cf.* p. 327], can produce a heat effect if they are brought in sufficient strength into a small space, as when they are allowed to fall upon a thermopile or other sensitive detector of heat.

Even the so-called chemical rays produce this same heat effect in a measure corresponding to the energy which is associated with them. Likewise, chemical action is not an especial characteristic of violet and ultra-violet rays. It was only by chance that the first substances to be decomposed by rays were those which are most readily decomposed by violet rays. At a later date substances have been found which are decomposed by yellow and red light.—The fact that we can see one ray and not another proves nothing at all. Even the ear can only perceive waves whose frequency lies within certain limits.

Almost all waves are perceived from the resonance caused when they are incident upon a suitable receiver. There is no organ, or even any other instrument, which can respond to waves whose frequencies vary over a range of more than fifty octaves. It is only those instruments, such as the thermopile or bolometer, in which resonance plays no part, but which convert all wave energy into thermal energy, that can receive and measure, without distinction, rays of all wave-lengths.

The heat rays which are emitted from a hot body are, in themselves, not hot. Heat is always first produced in the bodies which absorb the incident radiations. This fact is easily verified by means of the following experiment. A lens made of ice is readily constructed by freezing water contained in a suitably shaped dish. Light from the sun, which is ostensibly hot, can pass through such a lens without melting it, especially if all other rays except those of light have previously been absorbed from the sun's radiation. But if the radiation which has passed through the lens is allowed to come to a focus, any small quantity of powdered substance placed at that point may be raised to incandescence.

The differentiation of heat rays from light and actinic rays must not be allowed to drop as being unfounded, for we can distinguish between heat and light rays (let us say with the aid of our eyes) and between heat and actinic rays (perhaps by means of a photographic plate). But the difference is really determined by the properties of the receiving apparatus upon which the particular radiation is incident. All rays produce heating effects, which, in view of the laws of energy, only means that as long as there is a trace of energy present in the wave, this can always be converted into heat.

As heat rays, in the more strict interpretation of the term, we shall call those rays which are emitted by bodies in virtue of their high temperatures. From the standpoint of energy these are those rays whose energy is derived from the thermal energy of the radiating body. Although, as we shall soon see, bodies give out heat radiations at all temperatures, yet the intensity of these rays only becomes sufficient to be observed at high temperatures, when the wave-length is essentially smaller than that of electric waves. As regards wave-length, therefore, these rays occupy an adjacent position on the side of shorter wave-length in respect to the above-mentioned electric waves.

Heat Radiation and Temperature. At what temperature do bodies begin to radiate ? One might imagine that this would only happen at high temperatures since we do not notice any such radiation at room temperatures. But if we place a small vessel *L*, Fig. 154, containing liquid air, inside another, *B*, surrounded by ice-cold water, then bubbles of gas are seen to pass into the inverted tube *D*. This happens even if the space between *L* and *B* is free from all gases so that the heat can no longer be transmitted by their agency. When, however, the external vessel is filled with liquid air, at the same temperature as that in *L*, then the evaporation from *L* is stopped. If the external vessel contains liquid air at a lower temperature than that in L, then the temperature of this latter air falls. From this, and similar experiments, we have arrived at the conviction that bodies give out radiations at all temperatures. The emission of radiant energy is a process which only takes place in the radiating body independently of the surroundings, and which therefore only depends upon the properties of the source of radiation. But since all bodies radiate, it follows that every body must receive radiation from its surroundings. By emitting radiation a body loses energy ; when radiation is absorbed by a body energy is gained. The thermal equilibrium which exists between various bodies or different parts of the same body is therefore a dynamic one, produced by two processes counteracting each other. Every body is in equilibrium with its surroundings, as far as concerns the exchange of radiant energy, when it receives continuously the same amount of heat as it loses. If the external temperature increases, then it receives more radiant energy than it emits, in consequence of which its own temperature rises. This is followed by an increase in amount of its own radiation, until this becomes equal to the amount it is receiving. The condition of thermal equilibrium has then been attained. All bodies in the universe, including even the most distant stars, are therefore continually exchanging energy with one another.

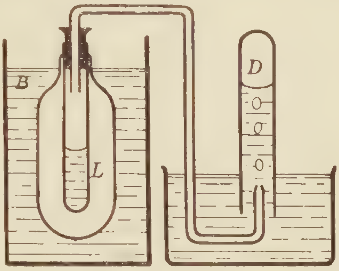


Fig. 154.

Wave-length and Temperature. The energy which a body radiates (its emissivity) is denoted by e. Even with the same body this quantity depends very largely on the temperature, and even at a definite temperature the emissivity varies with the wave-length of the radiation. When bodies, initially at room temperatures, are gradually heated, our eye does not at first detect any radiation, but a thermometer, or even our sense of feeling, will show that such a body is a source of radiation, and if the wave-length can be measured, it will be found that there is a preponderance of radiation having a long wave-length. If the temperature rises to about 500° C. the emission of red light from the hot body is perceived : suitable measurements will show the rays which were present at first are still there but their intensity has increased considerably. When the temperature is allowed to rise to 1000° C. the body appears yellow, at 1200° C. it is white, whilst at higher temperature the body is dazzlingly white. When the wave-length is measured it is found that as the temperature rises waves of shorter and shorter wave-length begin to appear.

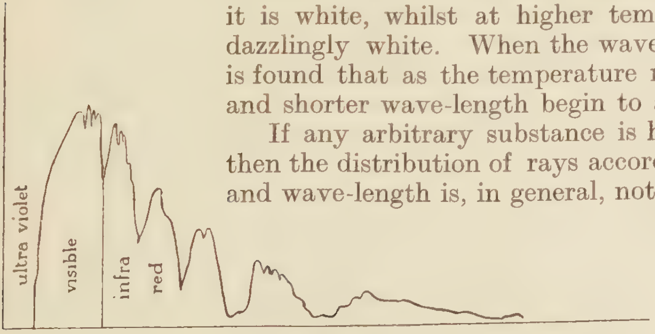


Fig. 155.

If any arbitrary substance is heated in this manner, then the distribution of rays according to their intensity and wave-length is, in general, not represented by a continuous curve ; the radiation corresponding to some wave-lengths is very feeble or even almost entirely absent. The distribution of intensity for the different radiations in sunlight is indicated in Fig. 155. The wave-lengths are plotted along the horizontal axis, whilst the intensities are represented by the ordinates of the curve. This selective radiation is peculiar to each individual substance. If, for example, a piece of platinum foil is partly blackened with indian ink and then heated so that it is red or even white hot, then the illumination is seen to be greatest over those parts which have been covered with ink.

Emission and Absorption. Still greater variations are shown by different bodies when we consider their behaviour with respect to the radiation they receive from external sources. A portion of the incident light is reflected at the surface of the body whilst another portion enters into the body. Of the portion which enters the body a part will be absorbed, the remainder appearing again on the other side of the body. Glass, for example, transmits a major portion of the incident light, so that it can be used as a window-pane ; but even so, some of the light is reflected by the surface of the glass so that we are able to use a piece of glass as a mirror. A smooth surface which reflects much of the incident radiation is said to be a mirror. Since the form of a wave is not changed by reflexion [*cf.* Fig. 43, p. 85], we are able to examine such sources as the sun by means of them. A matt or roughly polished surface reflects the incident radiation in random directions so that we are unable to learn anything more about the incident waves. If the light which is reflected from the surface of a body contains rays of all wave-lengths, then that body is said to be white ; when no radiation is reflected the body is called black. Black bodies have played a very important part in the development of the theory of radiation. Perfectly black bodies, *i.e.* those which do not reflect any radiation whatsoever, do not exist. Soot and platinum black (a finely divided precipitate of metallic platinum) are the most perfect of black bodies. If a closed hollow body is maintained at uniform temperature, let us say by immersing it in a heated bath (*e.g.* boiling water), then **Kirchhoff** has shown that the radiation which proceeds from a small opening in the surface of the body is the same as the radiation from an absolutely black body at this temperature ; because all the radiations which may be absent in the radiated energy are replaced by radiations reflected back from the hollow body. Such a hollow body therefore represents a perfectly black body as far as radiation is concerned, and, as such, has rendered much service in various researches.

The absorptive power of a body is the ratio of the absorbed to the incident radiation. This coefficient varies from one body to another and also depends upon the wave-length of the rays and the temperature of the body. From the facts which have already been stated, the absorptive power of a black body is unity, since all the incident radiation is absorbed. But because this ratio depends upon the wave-length, it follows that a body which is black with respect to one type of radiation is not necessarily black towards other radiations. If the value of a for all wave-lengths is unity, then the body is said to be absolutely black. The hollow body to which we have referred above is an absolutely black body, since the reasons why it is black are valid for all wave-lengths.

Kirchhoff's Law. Let e and a denote the emissive and absorptive powers of any one body, whilst E and A refer to the same values for a perfectly black body. Then $A = 1$. **G. Kirchhoff** (1824–1887) established the following law : if all these values are referred to the same wave-length and the same temperature, then for all bodies in general the ratio e/a is the same as if the bodies were all absolutely black. If therefore e_1, a_1 apply to one body, whilst e_2, a_2, e_3, a_3 . . . apply to others, whilst

E and A refer to an absolutely black body, then Kirchhoff's law states that

$$\frac{e_1}{a_1} = \frac{e_2}{a_2} = \frac{e_3}{a_3} = \ldots = \frac{E}{A} \quad . \quad . \quad . \quad (160)$$

But since $A = 1$ we may write $e/a = E$, or $e = aE$. The ratio of the emissive power to the absorptive power is, at any one temperature, and for any one wave-length, equal to the emissive power of a perfectly black body.

Since now all bodies which are not black have an absorptive power which is less than that of a black body, and therefore smaller than unity ($a < 1$), it follows that $e < E$, *i.e.* the radiation from a black body is a maximum. When therefore a curve is drawn representing the relation between the intensity of the radiation and its wave-length, for a black body, then the curves for all other bodies must lie below this ideal curve. They may perhaps touch the first curve at some particular points; for the corresponding wave-lengths the body acts like a black body. But the curves for a black body and any other body can never intersect one another.

The Laws of Black-body Radiation are therefore very important for all radiations. The following laws have been deduced with their aid.

The **Stefan-Boltzmann** law expresses the relationship which exists between the total radiation from a black body and its temperature. It states that this total radiation, W, is proportional to the fourth power of the absolute temperature, *i.e.*

$$W = CT^4 \quad . \quad . \quad (161)$$

where C is a constant.

Stefan discovered this law experimentally whilst working at Vienna in 1878; in 1884 it was derived by Boltzmann from theoretical considerations.

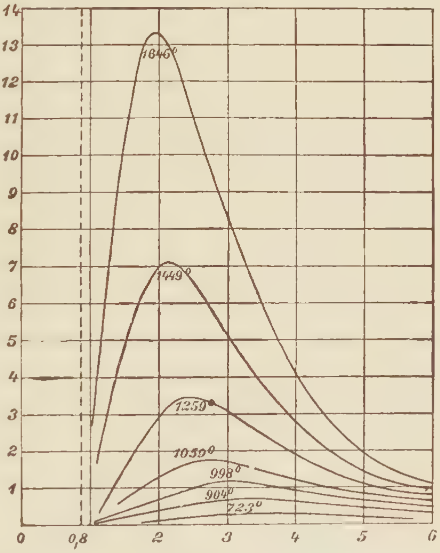


Fig. 156.

Some years later it was found possible to analyse the total radiation into its various constituents, so that the intensity of the different rays could be measured. It was found that there was always one wave-length, on either side of which the intensities of the rays decreased, so that at points farther removed from this particular wave-length the radiations became very feeble. In Fig. 156 curves are drawn in which the abscissæ represent wave-lengths, expressed in thousandths of a millimetre, whilst the ordinates represent the intensity of the rays for a perfectly black body. The lowest curve relates to a temperature of 728° K. If the experiments are repeated at higher temperatures similar curves, situated above the first one (723° K), are obtained. The most intense radiation, however, shifts gradually towards the direction of shorter wave-lengths. The higher the temperature, the greater does this shift become. In 1893 the following simple law was discovered by **W. Wien.** The wave-length of the most intense radiation is inversely proportional to the absolute temperature, so that the product of this wave-length and the absolute temperature is a constant. This law, known as **Wien's displacement law,** may therefore be written

$$T\lambda_{\max} = \text{const.} \qquad . \quad . \quad . \quad . \quad (162)$$

But even so, the chief problem concerning radiation was not yet solved. This obviously required that one should be able to state the wave length and intensity of the radiations emitted by a black body at any temperature. Many investigators had attempted to find this law. Experimentally the question had long been solved. The curves shown in Fig. 156 are the answer to it. It was now a matter of finding the equation to these curves, not on their own account, or for the mere pleasure of discovering such an equation, but because the exact equation could only be obtained when a conception of the processes at work in a radiating body had been properly realized. After many instructive attempts, which were partly successful in representing a portion of these curves, **Max. Planck** obtained the complete solution. In his equation λ denotes the wave-length of the various rays, whilst E denotes the intensity of the waves λ ; T is the absolute temperature, C and c are two constants, and e is the base of natural logarithms. Then the intensity E of a wave λ at a temperature T is expressed by the formula,

$$E = \frac{C}{\lambda^5} \frac{1}{e^{c/\lambda t} - 1} \qquad . \quad . \quad . \quad . \quad (163)$$

Since the equation is only true for black-body radiation, all questions concerning the effect of the actual material in the body do not arise ; the constants are quite general.

This formula has been well confirmed by many experiments covering the whole range of the spectrum. It gives values which are correct whether the wave-lengths are long or short, or whether the temperature is high or low.

The significance of this formula lies not so much in the fact that it represents the solution to a difficult and important problem, but rather in the notions and ideas which have led to its development. These conceptions have taken us far beyond the original problem and have directed researches concerning the structure of the atom into quite new and broad channels. It is therefore necessary for us to contemplate this theory in greater detail.

3. PLANCK'S QUANTUM THEORY

If the curves shown in Figs. 92 and 156 are compared with one another, a great similarity is discerned at once. The distribution of the radiant energy sent out by a hot body bears the same relation to its wave-lengths, as represents the distribution of molecular velocities in a gas. Is this similarity a mere accident ? Let us make a resumé of all the facts we know concerning the emission of this radiation. It is a thermal radiation, *i.e.* the energy of the radiation is supplied from the thermal energy of the radiating body. But since the energy of the individual molecules varies with their velocity, we may well suspect that the distribution of the molecular velocities will affect the way in which the radiation varies with the wave-length. Moreover, we know that the entities responsible for this radiation are electric in their nature; in fact, they are the electrons themselves. We shall therefore have to regard the electrons, which are always present in a body, as the source of the rays. They must be supplied with energy by coming into contact with the moving molecules; they then emit this energy in the form of radiation. Planck left the question as to the nature of these sources of radiant energy quite open. He only assumed that inside a body there were electric resonators, *i.e.* structures which could execute electric oscillations of a definite period, and which could be excited by molecular collisions.

These electric resonators produce electric waves in the surrounding space, exactly as other electrical waves are produced or, as in our problem, even the heat rays themselves. The different intensities which the observed heat waves have, arise therefore from the fact that the number of electrons which take part in producing particular wave-lengths is a variable quantity. We have therefore to determine this number. If an attempt is made to do this by using conceptions which have hitherto prevailed, due attention must be paid to what has already been said concerning the degrees of freedom in a body. According to the statements then made, the thermal energy is so distributed that each degree of freedom has the same amount of energy. Now the frequencies of the electrons must also be called degrees of freedom, since the energy of vibrations having one definite period can only be interchanged with that of others having the same period. The energy of a body whose period is given cannot be interchanged with that of a body whose period is different. When attempts were made to deduce, according to these laws, the distribution of energy in the spectrum of a hot body, impossible results were always obtained, since they contradicted those which had been established experimentally. There was always a tendency for the shorter wave-lengths to possess too much of the total energy.

After many prominent physicists had occupied themselves with this problem for many years, **M. Planck**, in 1900, arrived at the conclusion that a solution, in terms of prevailing ideas, was impossible, and that a fundamental change in our conceptions of Nature was necessary if the problem was to be solved at all. Planck maintained that energy could not be absorbed or emitted in arbitrary amounts but only in quantities which were fixed and definite. Whenever an exchange of energy occurred, the quantity involved was an integral multiple of these energy quanta. This was certainly an innovation. Hitherto energy had been defined as the product of a force and distance, magnitudes which we believe capable of changing by any amount however small. We cannot conceive why the velocity of a body could only change by definite amounts. Now an atomistic exchange of energy is not readily conceived, although we cannot quite regard it as an absurdity. But that is only part of the new theory. A further assumption made by Planck was that these quanta of energy were not all identical, but varied with the period of the electron concerned; in fact the quantum represents the smallest amount of energy which can be emitted in a wave whose frequency is ν. This amount is expressed by

$$e = h\nu \quad . \quad . \quad . \quad . \quad . \quad (164)$$

where h is a universal constant, having the same value for all frequencies, all substances, and all temperatures, whilst ν is the number of oscillations per second. It therefore follows that the quantum associated with rapid oscillations, *i.e.* when ν is larger, is greater than when the oscillations are executed more slowly. From the equation $e = h\nu$, we have

$$e = \frac{h}{T}, \text{ or } h = eT \quad . \quad . \quad . \quad . \quad (165)$$

where T is the period of oscillation. The following interpretation of this formula must be carefully avoided: h is *not* equal to the product of the energy e and the time in which this energy is radiated, for T is *not* the time in which the whole quantum of energy is sent out, but the time of one of the undetermined number of oscillations for which the indivisible quantum is emitted. The numerical value of this constant is very small; in fact

$$h = 6{\cdot}54 \times 10^{-27} \text{ erg. sec.} \qquad . \quad . \quad . \quad (166)$$

Now since the energy arises from the thermal movements of the molecules, it follows that the individual electrons will become capable of emitting electric waves when, from some chance collision, they receive an amount of energy of the magnitude of their quantum. For electrons having a high frequency, this quantum is much greater than for those whose frequency is lower; hence the chance of the former electrons receiving sufficient energy will be less, since these electrons have a greater frequency. It will only occur when a collision occurs with a molecule whose velocity is greater than the average velocity. But if the temperature of the gas rises, then the energy associated with all the collisions will increase, so that collisions which produce high frequency waves will become more frequent. The maximum intensity of the waves will therefore move towards the region of shorter wave-lengths. We therefore see quite clearly that according to this hypothesis of Planck a shift in the distribution of the radiated waves will take place, in such a sense that, at lower temperatures, the shorter waves will recur less frequently.

The more exact theory which Plank proceeded to derive from these new assumptions led to the above formula (163).

For what does it stand?

In the first place, the formula confirms the facts which have been observed in connexion with black-body radiation. According to the degree in which the results from these certainly most difficult experiments have become more precise, by the elimination of all possible errors, so does the agreement with Planck's formula become more and more perfect.

When the energies of the radiations of all the wave-lengths as given by Planck's formula are summed up, the unresolved total radiation of a black body is obtained: actually the result obtained is the **Stefan-Boltzmann** law.

If one attempts to find which is the wave-length whose intensity predominates, by investigating for which wave-length λ Planck's formula gives the highest value, the result obtained is exactly that expressed in Wien's displacement law, viz. that this wave-length is inversely proportional to the absolute temperature.

All these facts, however, cannot be regarded as new confirmations of Planck's theory, for they merely represent the original problems for the solution of which the formula was developed. If the formula had contradicted these laws, it would not have fulfilled the purpose for which it was instituted. A confirmation of the whole theory can only be found by applying the quantum theory to new problems which were not initially present. Such confirmations have been obtained in this instance in a manner which is almost without precedent.

Confirmations of Planck's Quantum Theory. In the first place, Planck himself was able to show from his more complete theory that the increase in the mean kinetic energy, k, of a single particle, when the temperature of a gas was raised 1° $C.$, could be obtained from the constants which appeared in his formula. Moreover the universal gas constant R, as we have already seen on p. 149, is the increase in energy per degree for one gram-molecule. If, therefore, the gram-molecule contains N molecules, then $N \cdot k = R$, so that we now have a new means whereby **Loschmidt's** number $N = R/k$ can be calculated. It was found to be in complete accord with the number obtained by the best of alternative methods, viz. $N = 60 \times 10^{22}$.

In 1905 **Einstein** extended the quantum theory, first developed by Planck, in order to explain the absorption and emission of heat radiation, so that it now applied to the energy of the atoms and molecules themselves. He assumed that an exchange of energy between two atoms could only occur in integral multiples of the quantum. We have already seen [*cf.* p. 158] in what way this line of thought has enabled us to explain, in particular, the specific heats of bodies at low temperatures.

Another problem which was likewise solved by Einstein with the aid of the quantum hypothesis concerns the phenomena to which reference has been made already [*cf.* p. 306]. It concerns the liberation of electrons from metals when they are under the influence of light. Einstein was able to use this new theory by assuming that the smallest amount of energy which could pass from the light to the metal was one quantum, $h\nu$, where ν was the frequency of the incident light. Each time this happened an electron was sent out from the metal. A portion of this quantum of energy, say P, was utilized in liberating the electron ; the remaining portion sufficed to impart a velocity to the electron. If this velocity is v, then the kinetic energy of the electron is $\frac{1}{2}mv^2$, so that Einstein's equation becomes

$$h\nu = P + mv^2/2 \quad . \quad . \quad . \quad . \quad (167)$$

The velocity v can be measured quite conveniently. For this purpose a positive potential is applied to the metal plate which is receiving the light radiation. The effect of this potential will tend to hold the electron in check, and when it has a certain value, the electron will just be prevented from leaving the metal. Now the electrons as they leave the plate carry with them a negative charge, so that the plate gradually acquires a positive charge. This positive charge will, of itself, increase until its value is sufficient to prevent any more electrons from leaving the plate. This limiting condition will manifest itself, for when the plate has acquired this definite positive potential, the process will cease. If V denotes this final potential, then according to equation (143), we have $V\varepsilon = mv^2/2$, so that the above equation becomes

$$h\nu = V\varepsilon + P \quad . \quad . \quad . \quad . \quad (168)$$

In this equation the frequency of the incident light is known from optical measurements, ε, the charge on an electron is also known, and V is obtained in the course of these experiments. It is therefore apparent that the value of h, Planck's constant, can be derived from this equation if P can be determined or eliminated from it. **Millikan** was the first to do this. He used two different sources of light, their frequencies being ν and ν' respectively. Since the work required to liberate the electron only depends on the nature of the metal, P must be the same in each equation. Hence, from the two equations :

$$h\nu = \varepsilon V + P \text{ and } h\nu' = \varepsilon V' + P$$

we have at once

$$h = \varepsilon\frac{V - V'}{\nu - \nu'} \quad . \quad . \quad . \quad . \quad (169)$$

P has vanished from this equation in which all the quantities except h are known. These experiments have given us the most reliable values for h; by using lithium and potassium, Millikan found $h = 6{\cdot}57 \times 10^{-27}$.

From this theory of the emission of electrons from metals when stimulated by light, several other deductions, each capable of a direct test, have been made.

(*a*) The liberation of electrons from metals only depends on the frequency of the incident light: if the quantum of energy associated with a light wave is not sufficient to set an electron free, then an electron will not be so liberated, however much the intensity of this light may be increased.

(*b*) But if light waves possessing larger quanta, *e.g.* violet or ultraviolet light, and finally X-rays, are allowed to fall on a specimen, it must be possible to liberate electrons from all bodies.

(*c*) The greater the frequency of the incident light, the greater is the energy which remains to accelerate the electrons after they have been set free. All these deductions have been amply confirmed by experiment. Conclusions have also been reached concerning the value of P. The electropositive metals are those which easily acquire a positive charge, *i.e.* it is not difficult to cause them to lose an electron. Consequently small quanta of energy are sufficient to liberate an electron from them. This too has been confirmed; in fact it was these alkali metals with which the phenomena of photoelectricity were first discovered.

4. LIGHT RAYS

The Spectra of Gases. Whereas the rays which are sent out by a hot body contain all wave-lengths from the shortest to the longest, in general, this is not so when the radiating bodies are gases. If, for example, sodium or one of its salts, such as sodium chloride, is introduced into a flame, that flame emits a strong yellow light. When the composition of this light is investigated, it is found to consist of one colour only, *i.e.* it is monochromatic. The rays which a sodium flame emits are therefore all of the same wave-length. By using refined apparatus it has been established that, strictly speaking, there are two wave-lengths, which only differ by a small amount from each other. At a later date it became possible to cause gases, even at ordinary temperatures, to emit light-radiations. In order to do this small tubes are filled with gases, the pressure of which is so low that an electric discharge can very easily be produced between two electrodes which have been sealed into the walls of the containing vessel. These are the so-called **Geissler** tubes, an older form of which is shown in Fig. 157. A more recent model is indicated in Fig. 158. From all gases contained in such tubes a large number of light waves of various wave-lengths are obtained. Even elements which exist in the solid state at room temperatures no longer emit a continuous spectrum when they are made gaseous by being placed inside the crater of an arc lamp. Instead, they emit a series of rays of definite wave-lengths, constituting what has been called a line spectrum. For example, in the iron spectrum, more than 1000 lines appear in the visible region of the spectrum. The

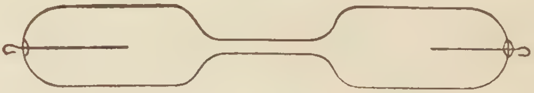


FIG. 157.

most noticeable thing about these lines, however, is the fact that the rays from any one element always have the same set of wave-lengths, wherever they may be produced. Finally, the technique of measuring these wave-lengths has been so far perfected, that we are able to measure them more accurately than anything else. It is now possible to distinguish between two light waves whose wave-lengths differ by less than one hundred-millionth part of 1 mm. (10^{-8} mm.).

For the first most important discoveries concerning the radiation emitted by gases and vapours, our thanks must be given to **Kirchhoff,** who, in conjunction with **Bunsen**, proposed these experiments in order to verify Kirchhoff's law, which we have previously stated [*cf.* page 334]. For gases, which are quite definitely not black-body radiators, since the light which they emit contains only well-defined wave-lengths, the general rule concerning the constancy of e/a is also true. It will be recalled that this fraction represents the ratio of the radiation emitted to that absorbed for rays of the same wave-length and at the same temperature. This also means that a gas, at a given temperature, will absorb rays of the same wave-length as those it can emit. By raising a piece of carbon, let us say, to incandescence, a spectrum can be obtained in which all wave-lengths exist. If this light is allowed to pass through an atmosphere of sodium vapour, then the light which is absorbed is just that which sodium would emit at the same temperature. Hence, at the exact point in the continuous spectrum where the yellow lines of sodium would be found, dark lines now appear. From the fact that a great number of these dark lines occurs in the spectrum of sunlight, we conclude that the sun is surrounded by an envelope of gaseous matter. The gases present in this atmosphere are those which have lines in their spectra corresponding to the wave-lengths now absent. A spectrum consisting of bright lines is known as an emission spectrum; that with the dark lines is an absorption spectrum. Absorption spectra give us just as much information concerning the path traversed by the light, as do emission spectra concerning the nature of the source of light. Spectrum analysis has been a very valuable means whereby elements could be recognized, a fact which is due in large measure to the small amount of substance necessary for this purpose. By its aid a whole series of previously unknown elements were revealed, so that the periodic system of elements became more complete.

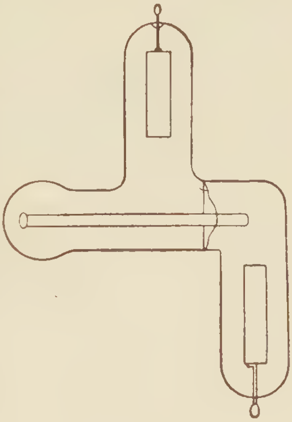


Fig. 158.

But spectra should be able to convey much more information to us. The reason for the unchanging length of the light waves can only be found in the definite properties of the elements. The spectrum with its numerous lines is like a handwriting in which the various elements are to be discerned again and again. He who can read it, is fortunate.

It may be thought possible that light rays would be like the various tones and overtones which a string produces; if so, the shorter wave-lengths would bear a simple numerical relationship to the longer ones. Even a rapid examination of a table of wave-lengths will show that this is not so.

Balmer's Formula. The first step towards the deciphering of this runic script was made by **Balmer,** a teacher at Basle. Without being guided by any theoretical considerations, he endeavoured to find a mathematical formula, which, by the mere changing of one of its terms, would give the various spectral lines. The element which he chose for this work was one in whose spectrum the lines were sufficiently numerous for a regularity to be indicated, but which were not so numerous as to cause confusion. He took hydrogen as being a suitable element, a choice which was very suitable, since it is not likely that his methods would have met with success had they been applied to the spectra of other elements. The formula which he discovered is as follows:

$$\nu_1 = K\left(\frac{1}{4} - \frac{1}{m^2}\right),\ m = 3, 4, 5 \ldots \qquad (170)$$

where K is a constant. This formula is somewhat altered from that originally given by Balmer, for instead of the wave-length λ, we have used the wave-number ν_1.* When the values 3, 4, 5 . . . were in turn given to m in this formula, he obtained the wave numbers corresponding to all the lines in the spectrum of hydrogen. The agreement which he obtained was not accidental; it is so perfect that not even the most refined and careful experiments which human skill has been able to devise have ever succeeded in detecting a discrepancy. By way of illustration a few wave numbers obtained experimentally together with the corresponding ones deduced theoretically are recorded in the following table.

TABLE 32

m.	$\nu \times 10^{-12}$ (calculated).	$\nu \times 10^{-12}$ (observed).
3	456·995	456·996
4	616·943	616·943
5	690·976	690·976
6	731·192	731·193
7	755·440	755·441
.	. . . .	. . . .
20	814·365	814·361

It was not possible to doubt but that the correct explanation of the spectra of gases would lead to Balmer's formula: this, however, would not be forthcoming from experiment, but by tracing back the phenomena to the original source of the radiation. This was achieved by the young Danish physicist **Niels Bohr** in 1913.

The Fundamental Postulates of Bohr's Theory.† Let us remind ourselves once again of the results already considered. There could be no doubt, especially in the case of gases whose molecules are so very independent of one another, that the starting point of the light rays must be sought for in the atoms or molecules themselves, in fact, in the electrical entities which constitute them. The atomic picture which **Rutherford** had proposed was an embodiment of all that was then known concerning atoms. The Balmer formula now indicated the direction in which further developments might be made. From the beginning, Bohr endeavoured to make use of Planck's quantum hypothesis in his attempts to increase our knowledge in this branch of physics. By combining these ideas with Rutherford's model of an atom he was enabled to derive Balmer's formula very simply. Bohr's theory was therefore not only a theory of light, but also an atomic theory at the same time. Hence any confirmations which might be forthcoming from experiments on light would also substantiate the properties of an atom which he had postulated in deducing his theory.

---

* The frequency ν is connected with the wave-length λ by the equation $\nu\lambda = c$, or $\nu = c/\lambda$. ν denotes the number of oscillations per second, or the number of wave-lengths in a distance $c = 3 \times 10^{10}$ cm. For the sake of greater convenience in calculations (because the factor c need not always be taken into account) another variable known as the wave number, ν_1, has been used. $\nu_1 = \frac{1}{\lambda}$, so that it specifies the number of waves per cm. Hence $\nu = c\nu_1$. As is usual in the literature of this subject we shall occasionally make use of one or the other of these two members. From the symbol ν or ν_1 it is clear which unit is meant.

† Although during recent years the original Bohr theory has been extended, and several important modifications made, it will still be more expedient to reproduce it in its earlier form at first, not only for the purpose of recording the historical develop-

At first Bohr made two fundamental assumptions :—

1. The electrons which, according to Rutherford's assumption, revolve around the nucleus of the atom, cannot move in any arbitrary orbit, but are constrained to move in one or other of a certain number of definite orbits. It is possible for an electron to revolve only in one of these paths. As long as an electron remains in the same orbit, the energy content of the atom is unaltered and no radiation is emitted from the atom ; the electron continues to revolve just like one of the planets.

2. But if some external agency produces a disturbance, the electron can move to another orbit. Each different orbit corresponds to a different energy content (potential and kinetic), so that the passage from one orbit or another means that energy is absorbed or emitted. When an electron moves to a more remote orbit, or, as we should say with respect to a terrestrial body, it is raised, energy must be supplied to the atom (a stone will not rise of its own accord). But when the electron passes from a higher to a lower orbit, energy is given up (just as a falling stone gives up its kinetic energy when it strikes the ground). Bohr assumed that the energy thus lost appears in the form of light energy, and so long as no additional disturbance arises, **the energy must appear as light energy ; he further supposed that the frequency of this light could be calculated from the quantum rule.** According to this $e = h\nu$. In our problem e denotes the energy liberated by an electron as it moves from one orbit to another so that the frequency of the light waves produced is $\nu = \frac{e}{h}$.

It is necessary for us to examine these two assumptions somewhat more closely. The orbits which are available for the motion of an electron around a nucleus, under the action of a force inversely proportional to the square of the distance, are those in which a certain relationship exists between the force of attraction and the velocity of the electron : this relationship requires that the attraction shall be equal to the centrifugal force, the lines of action of these forces passing through the centre of the atom. If Q is the charge on the nucleus, ε that on the electron, whilst r is their distance apart, then the force of attraction is $\frac{Q\varepsilon}{r^2}$. The centri-

ment of the subject, but also because in this first form the theory is most conveniently introduced. The more recent developments to which we shall refer in the final pages of this book will then be more easily understood and appreciated.

fugal force is $\frac{mv^2}{r} = mr \cdot \frac{4\pi^2}{T^2}$, where m is the mass of the electron. If we equate these two forces we obtain

$$\frac{Q\varepsilon}{m} = v^2 r = 4\pi^2 \frac{r^3}{T^2} \quad . \quad . \quad . \quad . \quad (171)$$

This is Kepler's third law concerning the motion of planets around a sun. We have proved that the electrons on their various orbits must have such a velocity that the square of their periodic time, T, is proportional to the cube of the radius of the orbit. Accordingly, all orbits would be permissible if the motion of electrons was only subject to the ordinary laws of mechanics. This freedom of motion, however, is limited by Bohr's first assumption, according to which the electron can only move in certain fixed paths. If the radius of the smallest orbit is a_1, then the radii of successive orbits are, according to Bohr, given by

$$a_2 = 2^2 a_1,\ a_3 = 3^2 a_1,\ a_4 = 4^2 a_1 \ . \ . \ . \ a_n = n^2 a_1,$$

i.e. the radii increase as the squares of the whole numbers.

The second postulate of Bohr concerns the differences in energy which arise in a system consisting of a nucleus and an electron, when the electron moves to different orbits. The mathematics of this problem has been dealt with already in the theory of potential. If a charge ε is to be brought from a point at a distance r_1 to another at r_2, work must be done if the motion takes place in a field of force. If Q denotes the charge on a sphere, then the work done, according to equation (104), is

$$W = Q\varepsilon\left(\frac{1}{r_1} - \frac{1}{r_2}\right),$$

and if the charge is brought up from an infinite distance, *i.e.* $r_2 = \infty$, the last term in the bracket disappears, so that we find the work which must be performed against the attractive forces of the system to be $W = Q\varepsilon/r$. The charge ε at a distance r is not at rest, however, but is revolving around the nucleus at a high speed. The revolving electron therefore possesses a high velocity, so that when the electron is shifted, part of its energy is utilized in doing some of the necessary work. How large is this portion ? It is just as big as the change in the kinetic energy of the revolving body. This result is easily obtained by using equation (171), which expresses the equality of the centrifugal and centripetal forces ; it is $\frac{mv^2}{2} = \frac{Q\varepsilon}{2r}$. Comparing this with the value $\frac{Q\varepsilon}{r}$ for the potential energy, the remarkable fact is observed that the kinetic energy of an electron is always equal to half its potential energy. The same result is of course true for planets moving in circles around a sun situated at their centre [*cf.* p. 59]. If we imagine that an electron has been attracted from an infinite distance, then the loss in potential energy is $\frac{Q\varepsilon}{r}$, which could be converted into some other form and given to external objects ; *e.g.* it might be converted into heat in overcoming resisting forces. But if the electron (or planet) is to revolve at a distance r from the nucleus, all this energy is not given up, for one-half must be utilized in increasing the velocity of the rotating body. Then, at a distance r, it just possesses that energy which it needs to revolve steadily and in equilibrium around the nucleus. It therefore follows that, when an electron (or planet) approaches the nucleus, only half its loss in potential energy can be given up, just as if, in order for the approach to take place, only half the amount of work has to be done when the body revolves, as when it moves from rest to rest. Hence, in removing an electron to infinity from an orbit of radius r_1, the work to be done is $\frac{Q\varepsilon}{2r_1}$; similarly for a radius r_2 it is $\frac{Q\varepsilon}{2r_2}$. If, therefore, the electron moves from an orbit r_1 to r_2 the amount of work necessary is

$$W = \frac{Q\varepsilon}{2}\left(\frac{1}{r_1} - \frac{1}{r_2}\right) \quad . \quad . \quad . \quad . \quad (172)$$

If the electron moves away from the nucleus, whilst if it moves towards the nucleus, *i.e.* from a higher to a lower orbit, this energy must be emitted from the atom.

Bohr's second assumption is that this energy which becomes available when an electron moves from a higher to a lower orbit is emitted as radiation, in fact, **the whole of this energy is equal to one quantum on each occasion.** The frequency of this radiation will therefore be such that its quantum is equal to the above value. Since the quantum of energy appropriate to a frequency ν is $h\nu$, we have

$$h\nu = \frac{Q\varepsilon}{2}\left(\frac{1}{r_1} - \frac{1}{r_2}\right),$$

so that the frequency of the radiation is

$$\nu = \frac{Q\varepsilon}{2h}\left(\frac{1}{r_1} - \frac{1}{r_2}\right) \quad . \quad . \quad . \quad . \quad (173)$$

This equation is really a mathematical representation of Bohr's second condition. Combining it with his first stipulation, that the radii of the possible orbits are not arbitrary but must be expressed by a_1, a_2, a_3 . . . we obtain for the frequency of the light emitted when an electron jumps from an orbit a_m to one a_n the formula

$$\nu = \frac{Q\varepsilon}{2ha_1}\left(\frac{1}{n^2} - \frac{1}{m^2}\right) \quad . \quad . \quad . \quad . \quad (174)$$

In this equation n and m can have all possible integral values subject only to the condition that $m > n$. If, for example, we put $n = 2$ and denote by R the constant magnitudes outside the brackets, we have

$$\nu = R\left(\frac{1}{4} - \frac{1}{m^2}\right), \text{ or } \nu_1 = \frac{R}{c}\left(\frac{1}{4} - \frac{1}{m^2}\right).$$

Putting $R/c = K$, the equation becomes

$$\nu_1 = K\left(\frac{1}{4} - \frac{1}{m^2}\right) \quad . \quad . \quad . \quad . \quad (175)$$

But this is just the formula which Balmer obtained, only with this difference : here it has not been obtained by trial, but by making definite assumptions concerning the nature and manner in which the waves are radiated. Since the two formulæ are identical, it is obvious that experiment will confirm Bohr's theory just as much as it does Balmer's formula. At present this is not to be wondered at, for Bohr's assumptions were made in order for the desired formula to be obtained. What is to be our final opinion of the theory we cannot say at present, for it is only an hypothesis. This will depend chiefly on whether the theory is able to explain problems which are at present puzzling, and whether Bohr's postulates will be confirmed by facts which were not used in developing the theory. Soon after its initiation, Bohr's theory was confirmed by so many facts that, for a time, its difficulties were not heeded. Before considering these confirmations, we shall find this a convenient place to view Bohr's fundamental postulate critically, so that we shall see at once the price at which the later brilliant results were obtained.

The Difficulties in Bohr's Fundamental Postulates. The first assumption made by Bohr was that, of the infinite number of orbits which classical mechanics offers to an electron, only a few are actually available. As to why the other orbits should be excluded, the theory says nothing. In addition, no reason is stated why an electron cannot approach a nucleus nearer than the first orbit, or why the energy which it then has cannot be given up, the electron thereby falling into the nucleus, towards which the force of attraction is pulling it.

Moreover, Bohr assumes that, as long as the electron remains in its orbit, no radiation is emitted. At first sight this assumption appears to be quite reasonable, and if one thinks only of the motions of the planets, it is so. The stability of their motions is due to the fact that no energy is expended by the system. It is not so, however, for a moving electron, because during its motion in space the æther must continuously become polarized and unpolarized, *i.e.* electromagnetic waves must be created. The assumption that the motion of an electron in a circular orbit is stable, is not in accord with the electromagnetic theory of light. Moreover, these same remarks can be applied to Rutherford's conception of an atom in which the electrons are assumed to revolve continuously. According to the electromagnetic theory, an electron which revolves continuously around a nucleus must send out electric waves whilst this motion lasts. In so doing, energy is lost, so that when all this has been expended, the electron falls into the nucleus. Bohr realized from the very beginning this conflict between Rutherford's model of an atom and the electromagnetic theory. But Rutherford's idea of an atom was supported by so many other experimental facts, that Bohr could not discard it ; instead, he gave up the electromagnetic theory and attempted to find some other explanation by which he could account for the emission of light. Let us now see what success he had.

Experimental Facts confirming Bohr's Theory. 1. A very excellent confirmation was obtained when the formula itself was subject to experimental test. Bohr not only presented equation (175) as we have done above, but he specified numerically the constants which appeared in it, and showed how they could be calculated from those already known, without introducing any further assumptions. In equation (174) the frequency ν would be determined, if we could only specify the value of a_1, the radius of the first orbit. At the same time this would make our picture of the atom more complete, since all the orbits in which it was possible for an electron to move would be known. This calculation was made by Bohr in several ways, but the same result was always obtained. The following proof is adduced on account of its simplicity. It depends upon the so-called **correspondence principle.** We have already observed, in discussing the quantum theory, that when quanta are introduced into an equation they do not make themselves felt so long as the quanta are small, *i.e.* when we are only dealing with slow vibrations (ν small).

The introduction of the quantum ideas did not appreciably affect the calculation of the distribution of the light rays in the various parts of the spectrum. Bohr now assumed that for small values of ν [*i.e.* n and m in equation (174) were large], the calculated frequency would agree with that deduced from the electromagnetic theory of light, *i.e.* from the period of rotation of the electron around its nucleus.

If we now calculate according to Bohr's theory the frequency, ν, of the waves produced when an electron passes from a very remote orbit $n + 1$ to the orbit n, we obtain from equation (174)

$$\nu = \frac{Q\varepsilon}{2ha_1}\left[\frac{1}{n^2} - \frac{1}{(n+1)^2}\right].$$

Now

$$\frac{1}{n^2} - \frac{1}{(n+1)^2} = \frac{2n+1}{n^2(n+1)^2}.$$

When n is large, unity may be neglected in comparison with it, so that the above expression reduces to $\frac{2}{n^3}$. Hence

$$\nu = \frac{Q\varepsilon}{ha_1 n^3} \quad . \quad . \quad . \quad . \quad . \quad (176)$$

The number of revolutions per second, z, made by an electron can be deduced from equation (171) by writing $2\pi rz$ instead of v, for $2\pi r$ is the distance traversed in one revolution, and if this is traversed z times per second then the velocity is $v = 2\pi rz$. Introducing this value into equation (171) and, at the same time putting $r = a_1 n^2$, we obtain

$$z^2 = \frac{Q\varepsilon}{4\pi^2 mr^3} = \frac{Q\varepsilon}{4\pi^2 m(a_1 n^2)^3} \quad . \quad . \quad . \quad (177)$$

Squaring (176), and then equating the result so obtained with (177), we have, since $\nu = z$,

$$a_1 = \frac{h^2}{4\pi^2 Q\varepsilon m} \quad . \quad . \quad . \quad . \quad (178)$$

This value, when inserted in equation (174), gives

$$\nu = \frac{2\pi^2 Q^2 \varepsilon^2 m}{h^3}\left[\frac{1}{n^2} - \frac{1}{m^2}\right] \quad . \quad . \quad . \quad (179)$$

In this formula it is only the value of Q which varies from one atom to atom. But since the charge on the nucleus is always equal to an integral multiple of the charge on an electron, we can put $Q = z\varepsilon$, where z is now the number of elementary charges in the nucleus of the atom. The above equation then becomes

$$\nu = \frac{2\pi^2\varepsilon^4 m}{h^3} z^2\left(\frac{1}{n^2} - \frac{1}{m^2}\right) = Rz^2\left(\frac{1}{n^2} - \frac{1}{m^2}\right) \quad . \quad . \quad (180)$$

where

$$R = \frac{2\pi^2\varepsilon^4 m}{h^3} \quad . \quad . \quad . \quad . \quad . \quad (181)$$

This value of the fixed term R can be calculated from the constants ε, m and h which are already known [ε is the charge on an electron, m is its mass, and h is Planck's constant]. For hydrogen $z = 1$. If, instead of using frequencies, we use wave numbers, then

$$\nu_1 = \frac{\nu}{c} = \frac{R}{c}\left(\frac{1}{n^2} - \frac{1}{m^2}\right),$$

so that $\frac{R}{c}$ must be equation to K, the constant in Balmer's formula.

1. When Bohr calculated the value of this constant it was found to be in excellent agreement with the value Balmer had found.

From observation the value of K in Balmer's formula = 109,723.

By calculation according to Bohr's formula $R/c = 109{,}000$.

A better agreement than this was not to be expected, since the values of ε, m, and h were not known with sufficient accuracy.

We are now able to calculate the radii of the different quantum orbits. In the case of hydrogen, we at present assume that there is only one electron moving around the nucleus; later on we shall discuss the reasons for this assumption. Then the charge Q on the nucleus is numerically equal to ε, so that we obtain from equation (178)

$$a_1 = 0{\cdot}532 \times 10^{-8} \text{ cm.} \quad . \quad . \quad . \quad . \quad (182)$$

This means that the *diameter* of the first orbit is approximately 1×10^{-8}, *i.e.* the hundred-millionth part of a centimetre [or 1 Å]. This result is also very satisfactory, since the order of magnitude of this diameter is the same as the diameter of the hydrogen atom deduced from the kinetic theory.

For heavier atoms the charge Q on the nucleus will be greater, in consequence of which the electrons will experience a greater force of attraction, so that the diameter of the first orbit will be less. We shall have to return to this point at a later stage.

Moreover, we can now calculate the velocity with which the electron revolves around the hydrogen nucleus, and also the number of revolutions per second. The value of v is obtained from equation (171) when, instead of writing r for the radius of the first orbit, we insert the value of a_1. The result is

$$v = 2187 \text{ km. per sec.}$$

The number of revolutions per second is deduced from $v = 2\pi rz$, and we find

$$z = 6530 \times 10^{12},$$

i.e. the electron revolves in its first quantum orbit around the hydrogen nucleus 6530 billion times per second.

2. Bohr's theory found a still more obvious series of confirmations, to which we shall now refer. At the time when these were being developed, they were very largely responsible for directing the attention of physicists to his work. From the beginning, the theory was not merely a derivation of Balmer's formula, but an important extension of it. In Balmer's formula, the denominator of the first term in the brackets is 4, *i.e.* $n = 2$. This means that the waves are produced when an electron jumps from the 3rd, 4th, or higher orbits into the second. Now according to Bohr's theory an electron should be able to pass into the first orbit, or from the 4th and higher orbits into the 3rd. From his formula the frequencies of the lines which would then be produced can be calculated. He found that for $n = 1$ the first line of the series appears in the ultra-violet, whilst for $n = 3$ it is in the infra-red. These regions in the spectrum of hydrogen were now investigated, and lines were discovered at precisely those places where theory had shown they were to be expected.

If, therefore, a hypothesis becomes confirmed when it has been found to possess a heuristic value, *i.e.* it allows us to make deductions which are afterwards established experimentally, then Bohr's theory has already been confirmed.

3. Finally, however, the number of lines appeared to be too numerous.

Two series of lines were discovered in the hydrogen spectrum which appeared to necessitate fractional values for n and m. This was contrary to the whole theory, which stipulated that n and m should be integral. The one series with $n = \frac{3}{2}$ was observed by **Fowler** in 1912, whilst the other with $(m + \frac{1}{2})$ instead of m was discovered in 1896 by **Pickering**, when investigating the spectra of certain stars. Bohr showed, however, that the formula can be so written that the fractional values disappear. To do this it is only necessary to note that

$$R\left[\frac{1}{\left(\frac{3}{2}\right)^2} - \frac{1}{\left(m + \frac{1}{2}\right)^2}\right] = 4R\left[\frac{1}{3^2} - \frac{1}{(2m+1)^2}\right] \quad . \quad (183)$$

The values for n and m are now integral, but the factor R has been increased fourfold. Now the factor K contains the term z^2 where z is the number of positive charges on the nucleus of the excited atom, so that if $z = 2$, instead of 1, the above apparent anomaly no longer exists. This, of course, means that the light waves do not originate from a hydrogen atom, but from some other element whose nuclear charge is 2. We have already learnt that 2 is the charge on an α-particle, the nucleus of a *He*-atom [*cf.* p. 315]. The difficulty will therefore disappear if it can be shown that the observed lines do not belong to hydrogen, but to helium. This is quite a possibility, since the lines observed by **Fowler** were found in hydrogen mixed with helium, this gas having been introduced because he believed that the excitation of the hydrogen spectra would be thereby facilitated. The experiments were therefore repeated using pure helium, all possible precautions having been taken to exclude hydrogen. The lines did appear in pure helium, their brightness being considerably increased. In this way, what had previously been an objection to Bohr's theory now supported it. This support was all the more welcome, since it was derived from an investigation of some other element than hydrogen.

4. A further important verification of Bohr's theory was obtained by **Franck** and **Hertz** at Göttingen in 1913. These investigators were very interested in questions concerning the calculation of the energy set free when a quantum is sent out, for it is obvious that an electron cannot keep on radiating energy unless its supply is replenished. When an electron has arrived at the innermost orbit it cannot radiate any more energy. Before this can happen it must be removed to an outer orbit, a process involving the expenditure of work in overcoming the attracting forces between the nucleus and the electron. The amount of this energy is exactly as large as that which is emitted as radiation in the reverse process, when the electron jumps into a nearer orbit. In both instances, apart from sign, it is $h\nu$, and since ν is known from Bohr's formula, this work can be calculated for each line in the spectrum. The energy difference corresponding to any two arbitrary values of n and m denotes the energy which is released in the passage from the mth to the nth orbit. When $m = \infty$ the second term disappears, so that the first term represents the work done when the electron is removed to an infinite distance from the nucleus. All connexion with the nucleus is thereby lost, *i.e.* the atom has become ionized. Franck and Hertz now conceived the idea of supplying this energy to a gas, by allowing cathode rays, to which energy had been imparted by means of an electric field in the manner already discussed [*cf.* p. 290], to pass through the gas. It will be recalled that if the difference in potential through which a cathode ray passes is V, then the

energy it acquires is $\frac{1}{2}mv^2 = V\varepsilon$. If a collision takes place between such an electron and an atom, this amount of energy will be given to the atom, so that it becomes possible for a light wave to be emitted whose frequency is ν, where

$$h\nu = V\varepsilon \quad . \quad . \quad . \quad . \quad . \quad (184)$$

When the amount of energy so supplied is sufficiently great, the electron will be removed to an infinite orbit, *i.e.* the atom will become ionized. From similar considerations the ionization potentials of different gases were obtained [*cf.* p. 305]. These considerations proved to be entirely conclusive, and the experiments which were designed by workers in many countries to test them revealed many new facts.

Still more remarkable were the results of Franck and Hertz when they endeavoured to excite each line of the spectrum in turn.

Hitherto, by the discharge of electricity through Geissler tubes or across the gap of an arc lamp, a whole series of lines had always been obtained. These two experimenters now sought to evoke each single line in the spectrum by suitably accelerating the cathode rays. The following limitation must be noted, however. In a beam of cathode rays, it is not possible to give to each single electron the same predetermined velocity; collisions occur between the electrons and gas molecules, so that they are retarded, their velocity therefore being less than the calculated value. There will always be present in such a beam, therefore, many electrons whose energy is less than that deduced from the potential difference, **but no electron can possess energy greater than this calculated value.** Hence, in the spectrum there will always be present lines whose quanta are less than that of the calculated value, but no line will be there whose energy quantum has a greater value than this.

The experiments were most successful when only those waves whose quanta were small were first investigated; the cathode rays were allowed to fall through such a difference of potential that only these radiations could be produced. When the potential difference was increased more lines appeared—these could then be easily distinguished from those produced at a lower potential.

In their first experiments, Franck and Hertz used mercury vapour. With this substance the conditions are very much simplified, since only monatomic particles are present. Electrons emitted from an electrically heated wire were allowed to pass through the vapour, their velocity being gradually increased by raising the potential difference. No lines were emitted until this difference had attained a value of 4·9 volts; the cathode rays simply passed through the gas without being appreciably weakened. But when the 4·9 volts value was reached something new occurred—in fact, a twofold effect was observed. In the first place, the cathode rays lost all their kinetic energy. It is true that before this stage had been reached collisions between electrons and atoms would have occurred, but the electrons had not given up any energy to the atoms, *i.e.* the collisions had been perfectly elastic ones. For when a small particle collides with a massive object—such an object being a mercury atom with a mass 200 relative to the electron of mass $\frac{1}{1800}$ – the former is reflected without any appreciable amount of its energy having been imparted to the massive body. But as soon as ever its velocity was greater than that corresponding to a potential difference of 4·9 volts, the collisions became inelastic at once, their entire energy being given to the atoms with which they collided.

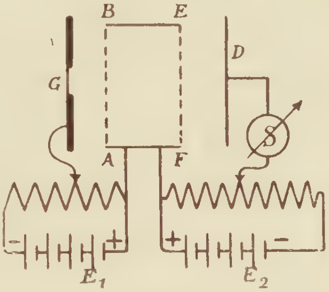


FIG. 159.

One of the experimental arrangements of Franck and Hertz is indicated in Fig. 159. Electrons were produced by heating a wire G electrically. This wire is surrounded by a grid AB [only one side of which is shown in the diagram]. A battery E_1 served to maintain this grid at a positive potential so that the electrons could be made to move with any desired velocity. Whilst moving with this velocity across the space between AB and EF, the electrons encountered the mercury vapour contained therein at a very low pressure. [There was no potential difference between AB and EF.] The plate D served to measure the end velocity of the electrons. When there was no potential difference between D and EF the electrons, after having traversed the vapour, arrived at D where they were discharged. The current which passed continuously from D was measured by means of a suitable detecting device S. When a small negative potential was applied to D the electrons were repelled somewhat, and when this had reached a value exactly equal, but opposite in sign, to that required to impart to the electrons their velocity, the electron current was reduced to zero. The potential necessary to reduce this current to zero was therefore a measure of the velocity of the electrons after having passed through the mercury vapour. The results of these experiments were as follows. So long as the potential between G and AB was less than 4·9 volts the potential to be applied to D in order for S not to be deflected was equal to that on AB. **The electrons had therefore passed through the mercury vapour without losing any of their energy.** But when the electrons were accelerated by a difference of 4·9 volts a new phenomenon appeared. **At this stage all the electrons lost their energy entirely.** The lowest possible potential on D was sufficient to reduce the current through S to zero. At the same time the mercury vapour **began to emit radiation.** Spectroscopic investigations showed that it consisted of one line in the ultra-violet region—its position was recorded photographically, and its wave-length found to be 2537 Å (0·00002537 cm.). Since then this line has often been measured. If the quantum of energy necessary, according to Bohr's theory, for this line to be emitted, viz. $h\nu$, is calculated, we have, since

$$h = 65{\cdot}5 \times 10^{-28} \text{ and } \nu = \frac{c}{\lambda} = \frac{3 \times 10^{10}}{2537 \times 10^{-10}},$$

$$h\nu = 7{\cdot}8 \times 10^{-12} \text{ erg.}$$

The energy appropriate to an electron having fallen through a potential difference of 4·9 volts is, according to equation (143) $= V\varepsilon$. Now $V = 4{\cdot}9/300$ *E.S.U.*, and $\varepsilon = 4{\cdot}77 \times 10^{-10}$, so that $V\varepsilon = 7{\cdot}8 \times 10^{-12}$ erg. From the confirmation gained from these two results, it is seen that the emission of radiation only occurs when the energy of each colliding electron is sufficiently large to excite the first line in the mercury spectrum.

When the potential on AB was still further increased beyond 4·9 volts, the potential on D necessary to cause no current through S was always 4·9 volts lower than that on AB. This means that the electrons, in passing through the mercury vapour, lose just that amount of energy which corresponds to a fall in potential of 4·9 volts. But when the residue of the original potential, after reduction by 4·9 volts, was again 4·9 volts or a multiple of this, the electrons lost this higher velocity of 2 × 4·9 volts, 3 × 4·9 volts, etc., also.

From these observations we conclude that mercury vapour cannot absorb any arbitrarily small quantity of energy. So long as the velocity of the electrons is below 4·9 volts, none of their energy is imparted to the vapour, but immediately this value is reached the whole energy content of the electron is given to the mercury atoms. In this phenomenon we have a direct proof that the energy of a gaseous molecule can only be increased by definite small amounts, as Bohr's first assumption postulates. Since then many other similar experiments have been made. They deal chiefly with the monatomic inert gases and metallic vapours, and in each instance this same definite assertion can always be made.

The Emission of Radiation from Excited Atoms. If we again attempt to visualize the process by which light is sent out according to Bohr's theory, then two parts are clearly discernible. Let us think of a hydrogen nucleus and its one electron. In general the electron will be in the orbit nearest to the nucleus. The potential energy is then a minimum and the atom is absolutely incapable of emitting further energy, just like a stone resting on the ground. If the electron is to emit radiation, energy must first of all be imparted to it. This can only occur if the electron is associated with some external influence ; just as the stone must be lifted from the ground, so the electron must be raised to a higher orbit. When this takes place the atom is said to be excited. It may be that the atom can remain excited for a certain small time, *i.e.* the electron will continue to revolve in a higher orbit without having to emit radiation. Attempts have even been made to estimate this time—it lies between 10^{-8} and 10^{-9} sec. **Mie** and other workers believe that the electron returns at once to the innermost orbit and that the time measured is that in which the radiation is emitted. This last assumption seems to be the more satisfying, since we do not readily see what is the effective cause which keeps the electron for a little while on a higher orbit and then allows a quantum jump to take place.

An electron can return from an outer orbit to the innermost orbit in several ways. It can leap across all the intermediate orbits into the first. The radiation which is thereby emitted is represented by $n = 1$, $m = \infty$ in Bohr's formula ; it is the shortest wave-length radiation possible for this particular atom. An electron can, however, pass into one or more intermediate orbits before reaching the first, and for each jump the radiation emitted has a lower frequency.

An atom must therefore be excited until it is completely ionized, before the whole spectrum can be emitted.

Bohr's formula always gives the frequency as the difference of two terms, each of which denotes the limiting wave which is sent out when the electron jumps from a great distance into the nth or the mth orbit. If these limiting values have been ascertained for the arbitrary values $n = 1, 2, 3, 4, \ldots$ then all the possible differences between any two of these limiting values is a possible frequency. It is easily seen that it is not very difficult to account for the great number of lines present in the spectra of many elements.

A very important question which remains unanswered concerns the means whereby an atom can become excited, and the source from which the atom receives energy after radiation has been sent out. In general, all agents capable of putting a restraint upon an atom are capable of exciting it. It is always a question of disturbing the normal state of the atom by some powerful agent for an instant, after which the atom returns to this state, and in so doing radiation is emitted.

1. Atoms may be excited by bombarding them with cathode rays. The way in which atoms are thereby excited has already been discussed. This method has the advantage that the increase in energy can be altered at will, and even measured [*cf.* pp. 349–352].

2. As a second means of producing this effect, reference must be made to high temperatures, when the molecules of a gas are in a state of violent agitation. It is therefore well within the range of possibility that an atom may collide with another in such a way that its electron moves to a more remote orbit. We have to consider whether or not the exchange of energy which takes place during a collision is sufficient for the gas to emit its characteristic radiation. Since we are able to specify the quantum of energy associated with all waves, and, moreover, since we know the velocity and thermal energy of gaseous molecules, the above question can be answered at once. Using equation (92) [p. 212] the values of the kinetic energy of a molecule at different temperatures are as follows:

TABLE 33

T° (abs.).	500.	1000.	2000.	5000.	10,000.
Kinetic Energy (ergs × 10^{13}) .	1·03	2·06	4·12	10·3	20·6

Besides, from a knowledge of Planck's constant h, and the frequency ν, we can calculate the energy $e = h\nu$, which is necessary in order for a wave of this frequency to be emitted. These values have already been given in Table 30 [p. 327] for the whole range of electromagnetic waves.

If we compare the molecular energy with the energy quanta in this table we see at once those wave-lengths which can be excited at various temperatures. Attention must now be called to another point, however. The calculated thermal energy is really a mean value, for in every gas there are present a considerable number of molecules moving with velocities which are greater than the mean velocity, in fact, some of them have velocities considerably in excess of this value. With the enormous number of molecules present, say in a flame, this effect becomes noticeable.

Moreover, in calculating the amount of energy exchanged during a collision between two molecules, we must remember that it is not the absolute but their relative velocity which must be considered. If the two molecules make a head-on collision this relative velocity is twice as great as the velocity of each individual.

Further, amongst all the milliards of encounters which a molecule makes in a second, there is no evidence to show that an electron, after it has been moved from the first to the second orbit, say, shall not be removed to a more remote orbit by another collision before sufficient time has elapsed for radiation to occur. Afterwards energy can be sent out which will correspond to two or more collisions.

If due consideration is paid to all these points, it is found that the energy transferred during a molecular collision at fairly high temperatures may be sufficient to excite a red or infra-red line, but shorter wavelengths than these are only possible at exceptionally elevated temperatures.

Even so, the essential point in the problem has not yet been discussed. In the spectrum of every gas, lines are always present in the red and infra-red region, *i.e.* the energy quanta are small. Now such radiation is emitted when an electron jumps from some remote orbit into one which is still far from the nucleus, let us say, from the 25th to the 20th. But it must be remembered that, in the unexcited state of an atom, the electrons are always revolving in the innermost orbit and that radiation can only take place **when the electron has been raised from this orbit**. Thus the excitation of the line 25 to 20 would be by no means ensured by supplying the electron with the energy corresponding to the difference of these levels. The electron is normally only in the first orbit and not in the 20th, to which position it must first be raised. The least amount of work which is sufficient to excite an atom must therefore be capable of shifting the electron from the first to the second orbit, and this amount must be supplied before anything happens to the atom. In a somewhat restricted sense this is called the energy of excitation, and it has been measured in the case of many elements.

If this value is compared with the thermal energy of a gas, it is still doubtful whether in the more usual types of flame the energy is sufficient to excite an atom, if we only consider the mean temperature of the flame as ordinarily measured.

Let us now return to the consideration of some experimental data. Ever since the discovery of spectrum analysis by Bunsen and Kirchhoff, it has been known that the alkali metals can very easily be caused to emit their spectrum—it is only necessary to place them in a hot flame. Even in the flame from methylated spirit the well-known yellow line of sodium and the red line of rubidium can be excited ; but these are the lines corresponding to the passage of an electron from the first to the second orbit. As the temperature is gradually raised, fresh lines make their appearance one by one. In a very hot bunsen flame, about 1550° *C.*, **Franck** found for lithium 1, sodium 2, potassium 3, rubidium 4, and cæsium 6 lines. Cæsium has the lowest excitation energy—it is $2{\cdot}2 \times 10^{-12}$ erg. An objection can always be urged against these experiments on the grounds of the chemical changes (combustion) which are always taking place in a flame, so that it is not yet possible to state definitely whether or not temperature alone is sufficient to excite an atom. At the very high temperatures known to exist in the hottest stars, there can be no doubt that the temperatures there are sufficient to cause an atom to emit radiation. If this is so, then we can readily comprehend why the easily ionized elements such as cæsium and rubidium were not at first detected in the sun. At the high temperatures there present, these elements would be completely ionized, *i.e.* they would have lost their outer electron, which is responsible for the emission of light. Afterwards, however, by examining the coldest portions of the sun's surface, the sun-spots, these elements were discovered.

3. A third agent which is capable of exciting atoms is light itself. According to the quantum theory, bodies should receive energy in quanta from radiation incident upon them, these quanta depending upon the frequency of the light. The phenomena of photoelectricity have already taught us how this hypothesis has been confirmed by experiment. In such a process, an electron is completely removed from its parent atom, and we have learnt that an electron is most readily liberated from the

alkali metals. If now such radiation is allowed to pass through a gas, it must ionize the gas if its quanta are sufficiently big ; even if the quanta are not so large, the electron will be raised to one of the higher orbits. Moreover, it is essential that the light quanta should be just as large as those which are afterwards emitted ; in other words, the gas must be irradiated with light of the same frequency as it emits. For example, when sodium light traverses sodium vapour, the quanta which it receives are just sufficient to raise the electron from the first to the second orbit, so that the atom is then capable of emitting light of this same frequency. These yellow sodium lines are known as resonance lines.

If the light quanta are greater than this, the electrons can be raised to higher orbits, and the return to the first orbit can take place in one or two stages. Should the return occur in one step only, then the frequency of the light emitted is a maximum, and equal to that of the incident radiation. If the return takes place in several steps, however, then light is given out after each jump, but the quantum of energy for this radiation is only a fraction of that for the incident light ; this means that the frequency is less, or that the colour of the light has shifted towards the red end of the spectrum. This light, excited by some incident light, is called fluroescent light. The above remarks provide an explanation of the well-known fact that the fluorescent light, with respect to the incident light, has a wave-length nearer to the red end of the spectrum. In the limiting case the wave-length of the emitted light can be equal to that of the incident radiation : in such a case we speak of **resonant fluorescence.** This word had been in use for a long time, because the conception had already been formed that we were dealing with a phenomenon in which the equality of two frequencies was to play a decisive rôle.

Photochemical actions also depend upon the absorption of energy from the incident radiation. By the term photochemical action is understood chemical changes which are produced by the action of light. According to the exhaustive researches of **Franck**, in such actions, radiant energy is first absorbed from the radiation emitted by the source ; this is not emitted again, but converted into kinetic energy, through which the chemical process is initiated. In connexion with this, **Warburg** discovered the simple law that the number of transformed atoms is equal to the number of quanta absorbed.

The Range over which Bohr's Formula is Valid. When the formula (174) was deduced, it was assumed that we had only to contemplate one electron revolving around a nucleus. If we have not mentioned this fact expressly, we have never considered the effect of any other electrons on that which was under discussion. The formula is therefore applicable in the first place to the hydrogen atom, which consists of a nucleus with only one electron. But as soon as several electrons revolve around the nucleus, they will of necessity influence one another, so that the above formula, as we have presented it, is no longer applicable. In the helium atom we have a nucleus with two electrons around it, but the influence of one electron on the other will not be very pronounced. Actually, the formula holds very well when applied to the helium atom. But if we remove one electron from a helium atom we have a positively charged helium ion, with one electron revolving around the nucleus of charge $z = 2$. The formula can therefore be applied, and it is found to agree with the experimental data.

For mathematical reasons it is impossible to deduce a formula which shall be strictly applicable for a nucleus with two attendant electrons. The so-called three-body problem has been tackled by many astronomers, but so far no solution in finite terms has been forthcoming. The chance of solving the problem when there are 10, 20, 50, or even 90 particles all exerting an influence on one another is very much less. It does not follow, however, that the whole idea of Bohr's theory is untenable in such instances. It is certainly not so. Even then we have a series of predescribed orbits in which the electrons move, and by the passage of an electron from one orbit to an inner one the energy liberated will be radiated away. But this energy can no longer be calculated in a simple manner from the radii of these orbits, since the number and positions of the remaining electrons have to be considered.

For such cases, however, **Ritz's combination principle** has been found very useful. This principle states the following : if the energy necessary to remove an electron from the various orbits to infinity has been found, *i.e.* one knows the values of the various terms, then the difference between any two terms is always equal to the frequency of some line in the spectrum.

The disturbing effect due to neighbouring electrons will be greatest when the electron under discussion is surrounded by others. But if the most remote electron is raised to a still higher orbit, so that it encircles the remaining cloud of electrons, then the effect of the nucleus and these remaining electrons will be as if they were all situated at the centre of the atom. If the nucleus has a positive charge z (expressed as an integral multiple of the unit positive charge ε), then there will be z electrons in a normal atom. Let us consider one of these outside all the others. It will be attracted by the charge $+z\varepsilon$ towards the nucleus and will be repelled by the $(z-1)$ electrons, the total charge on which is $-(z-1)\varepsilon$. The excess positive charge is therefore ε, which is numerically the same as that on the nucleus of the hydrogen atom. Consequently the motion of this outermost electron will be represented very closely by the simple hydrogen formula. As we shall see in the next section, this is especially so for the alkali metals (Na, K, . . .). The electron is able to return from a very remote position in one or several jumps to the innermost possible orbit, but since all these jumps take place between remote orbits the exchange of energy will be much less, from which it follows that the frequency will be less, and the waves will lie well within the red and infra-red regions.

Elliptical Orbits. Since the attraction between nucleus and electron, like that between two planets, varies inversely as the square of the distance, the electrons can move not only in circular but also in elliptical orbits. But the elliptical orbits are likewise governed by the rules of the quantum theory. The major axes of all possible ellipses belonging to one shell of electrons are all equal, in fact they are equal to the diameter of the circular orbit for that shell. The minor axis is always a simple fractional part of the major axis. If, for example, the major axis is 3, then the minor axis can be only 1, 2 or 3. These orbits are denoted by the symbols 3_1, 3_2, 3_3, the suffix specifying the minor axis. For these elliptical orbits the nucleus is at a focus, so that for a short time the electron approaches the nucleus more closely than do the electrons of the second circular orbit. It is seen therefore that one can no longer speak of real shells, for the word now only denotes energy levels. For the second shell only two orbits are possible ; their axial ratios are 2 : 2 and 2 : 1, *i.e.* a circle and an ellipse, whilst for the first shell there is only room for a 1 : 1 orbit, *i.e.* a circle.

These varied and different elliptical orbits for the more remote shells not only occur with elements having many extra-nuclear electrons; they are a possibility even in the simple hydrogen atom, for when the electron has been removed to a remote level it can revolve in a circle or in one of the possible elliptical orbits. In Fig. 160 all the possible orbits belonging to the first four shells for a hydrogen atom have been drawn. They are the orbits 4_4, 4_3, 4_2, 4_1, then the 3_3, 3_2, 3_1, 2_2, 2_1, and finally the 1_1 orbit. The reason why the orbits have to be drawn in this way is to be found in the fact that certain minor characteristics were discovered in the radiation emitted; we refer to the so-called fine structure of spectral lines, which is only to be observed with the aid of the most perfect optical instruments, but which has most certainly been shown to have a real existence. By introducing this idea of elliptical orbits **Sommerfeld** was able to account for these lines. When a planet is moving in an elliptical orbit, its velocity is greater when it is nearer to the sun (perihelion) than when it is further away (aphelion). But when the velocity of an electron changes, its mass alters [*cf.* p. 298]. Sommerfeld showed that this change in the mass of an electron would produce a slow rotation of the major axis of the ellipse [*cf.* Fig. 161]. After 40,000 or more revolutions around an elliptical orbit the major axis will have made one complete rotation. It therefore follows that the energy content of a particular orbit will be influenced by the measure of its minor axis, so that not quite the same wave is emitted when an electron jumps into a 3_3 orbit as when it moves

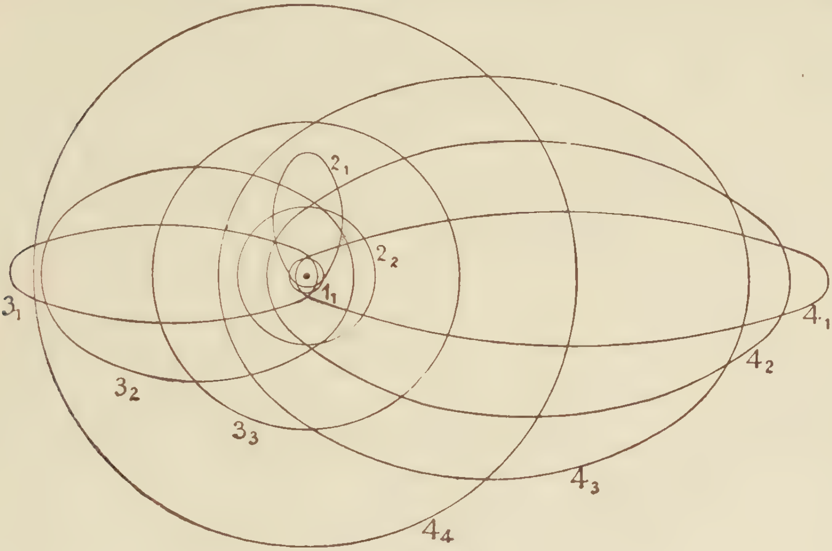


Fig. 160.

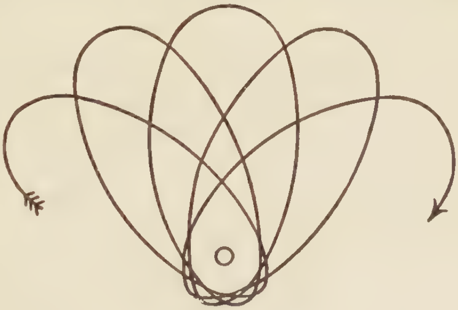


Fig. 161.

into a 3_2 orbit. These small differences were responsible for the fine structure of the lines.

5. X-RAYS

Another region of the spectrum for which Bohr's theory holds, is that in which X-rays occur. We have already learnt that X-rays are produced whenever rapidly moving electrons fall on solid bodies. In order to produce X-rays of maximum intensity, the cathode is made somewhat concave, so that the electrons sent out from all points of the cathode *A*, Fig. 162, shall be brought to a focus. The heat produced at this point is so great that if it were allowed to form on the walls of the tube the glass would either melt or crack. The curvature of the cathode *C*, however, is such that the cathode rays are focussed on to a metal *A*, known as the anticathode; very often this is cooled artificially. The X-rays which proceed from this body now possess the advantage of being able to produce a well-defined shadow of any object, such as a limb of our body, placed in their path. This is not possible unless the cathode rays have been sharply focussed on *A*.

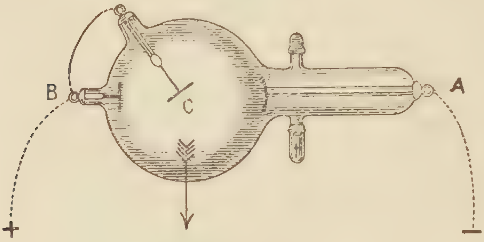


Fig. 162.

In these days X-rays are generally produced by means of a tube which was originally designed by **Coolidge**; a modern form of such a tube is shown in Fig. 163. *F* is a tungsten filament heated by means of an electric current to such a temperature that electrons are emitted. *A* is the anticathode and has a very high positive potential with respect to *F* [actually *A* is earthed and *F* is connected to the secondary of a transformer]. In consequence of this potential difference the electrons acquire a very great velocity, so that when they fall upon the anticathode Röntgen rays are emitted. The great advantage of this form of tube is that the number of electrons in the cathode beam and their velocity can be altered independently of one another.

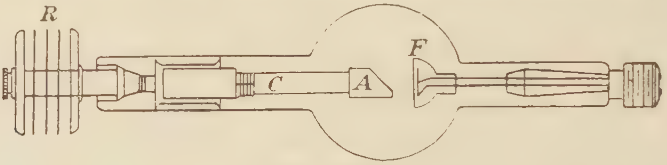


Fig. 163.

Means of detecting X-rays. We are enabled to detect the presence of X-radiation from the fact that it causes a fluorescent screen to luminesce or renders a photographic plate black, after subsequent development. The first method gives us an instantaneous picture, whereas the second provides us with a permanent record. For measuring the intensity of these rays use is made of the fact that they are able to produce ionization in a gas. The measurement depends upon the production of a saturation current in an ionization chamber [*cf.* p. 303].

The Nature of X-rays. For more than ten years after their discovery the nature of Röntgen rays had not been revealed. Although they had been used in medicine and the technique greatly improved, nobody knew what really constituted X-rays. All attempts to deflect them by means of electric or magnetic fields had proved unsuccessful. They were not like alpha or beta rays, since no charge was associated with them. They were therefore regarded as æthereal oscillations akin to light. More experiments were then carried out by which it was hoped to show that interference, diffraction and polarization could be obtained with X-rays. For a long time no positive results were obtained. But **Marx** had shown that they were propagated with the velocity of light. A few somewhat uncertain indications on photographic plates pointed to the fact that their wave-length was about 4×10^{-8} cm. All these experiments tended to strengthen the idea, which was gradually becoming prevalent, that in X-rays we are really dealing with light radiation of very small wave-length, perhaps a thousand times less than that of visible light [*cf.* Table 30, p. 327]. This was completely in accord with the fact that no success had been forthcoming when attempts had been made to measure their wave-length. For in all such experiments the wave-length must be comparable with the dimensions of the apparatus employed. When the wave-length of light is measured, the distance between lines ruled upon a glass plate can be used for purposes of making the comparison. The finest of these so-called gratings contain about 5000 lines per cm., all of them being at exactly the same distance from one another. If X-rays had a wave-length which was quite as small as that conjectured, then such gratings would be far too coarse, and it is not within the range of human technique to produce gratings with a greater number of lines per cm. In 1912 **Laue** imagined that Nature herself would provide in crystals gratings which were sufficiently delicate. Since crystals are forms endowed with great regularity, it might be that the atoms in them are arranged one after another according to some very definite and fixed plan. It is quite easy to calculate the distance between neighbouring atoms in such structures. A knowledge of their density gives us the mass per unit volume, whilst from the atomic weights and **Loschmidt's** number we can calculate the mass of the individual atoms or molecules in the crystal: *e.g.* we know the atomic weights of sodium and chlorine which unite to form crystals of sodium chloride. Consequently the number of sodium chloride molecules per unit volume can be obtained by dividing the density by the mass of a single molecule. In this way we know the volume, assumed to be cube, associated with one molecule. The edge of this cube is the distance between two adjacent particles, and constitutes our new measuring rod. The most careful measurements give $2{\cdot}8 \times 10^{-8}$ cm. as the distance between two atoms in a crystal of sodium chloride. Using such crystals as gratings, and X-rays in place of light rays, **Friedrich** and **Knipping** succeeded in obtaining diffraction spots upon a photographic plate. In order not to interrupt our train of thought, the details of their experimental arrangement will not be discussed. We shall be able to return to that subject in connexion with our consideration of similar problems concerning light waves. For the moment it is sufficient for us to know that it was now possible to measure the wave-length of X-radiation, although it was less than the ten-millionth part of a centimetre. Many such measurements have since been made. Similar advances to those which had been made in the study of light waves and radioactive bodies were now made with X-rays. Apparatus for producing high voltages had already been perfected; since it was at the immediate disposal of X-ray workers, our knowledge of this subject increased very rapidly.

The X-rays which proceed from the anticathode of an X-ray tube contain rays of very different wave-lengths. In the first place the radiation consists of two portions—one is the continuous radiation like the heat radiation from a solid body only its wave-length is very much shorter. The lengths of the waves in this radiation vary continuously from one part of the spectrum to the next. But in addition to this radiation there is a line spectrum which stands out very clearly from the other radiations—it is similar to the radiation proceeding from a Geissler tube containing a gas. In an X-ray line spectrum only a small number of lines, each having a definite wave-length, are present.

The Continuous or Impulse Spectrum. When the intensity of the different waves is investigated in this portion of the spectrum, as for thermal radiations, the following results are obtained. Towards the region of longer wave-lengths the curve passes slowly downwards, in much the same manner as for heat radiations [*cf.* p. 335]. At the other end of the curve a maximum is reached, after which the curve drops suddenly to zero, a fact which shows quite clearly that there is an upper limit to the frequency of the waves present in this continuous radiation. If X-rays are really an electromagnetic disturbance, then the fundamental facts of the quantum theory must be applicable to them. Now the quanta necessary for waves of such high frequencies to be emitted will be much greater than those associated with the emission of visible light. From the fact that X-rays are produced when cathode rays bombard the anticathode, one is led to believe that it is these rays which will supply the necessary energy quanta to the atom. We have already seen that cathode rays can acquire any given amount of energy by allowing them to pass through a suitable potential difference. But since many of the cathode rays will collide with other particles present in the field, their velocity will be diminished accordingly. Hence, in a given stream of electrons, there are always some whose velocity is a maximum, which will depend on the potential difference through which they have passed; the velocities of the others will be distributed irregularly below this value. The way in which the impulse spectrum is produced is supposed to be somewhat as follows: the cathode rays penetrate into the metal anticathode, where they experience considerable retardation. This change in velocity causes a disturbance in the electromagnetic field around them, which is propagated as an X-ray. If this is so, then the X-rays having the highest quanta must be due to those electrons which were moving most quickly. In 1907, **Einstein's** law concerning the photoelectric effect, viz. $h\nu = V\varepsilon$, was interpreted in the reverse direction by **Wien** and **Stark.** They realized that electrons falling upon matter would produce electromagnetic waves whose frequency ν would be such that $h\nu$ would be equal to the energy $V\varepsilon$ of the incident electrons. We have already seen [*cf.* p. 349] how this law was verified in the case of light waves by the experiments of **Franck** and **Hertz.** An opportunity now presented itself of testing the law for the shortest waves in the continuous X-ray spectrum. The experiments of **Wagner** (1918), in which he measured first the frequency of these waves and then the exact potential difference through which the cathode rays had passed, provided ample confirmation for this formula when it was extended to include Röntgen rays. Both these quantities were measured with precision, so that from these experiments it was possible to calculate h very exactly. For ε was accurately known from **Millikan's** work, and since ν and V were now known, the only term remaining unknown in Einstein's equation was h; it could therefore be calculated.

The confirmation provided by these experiments must not be interpreted to mean that it had been established that the energy of the X-rays was equal to that of the cathode rays. Such a proof could only be forthcoming from experiments in which the total energy of both the cathode rays and the X-rays which they produced was measured. This had not been attempted in the above experiments. Other workers have shown, however, that only 0·2 per cent. of the energy of the cathode rays becomes available in the form of X-ray energy. The rest is converted into other forms of energy. Moreover, we do not know whether all the energy of one electron becomes converted into one X-ray quanta. But if only one quantum of radiation can be produced by one electron, **then the greatest quanta, $h\nu$, are those which originate from the most rapid electrons, if all their energy is utilized for the production of the quantum.** And this is what equation (184) expresses.

The most accurate measurements have been made in the region between 4500 and 60,000 volts. At the present time, higher potentials than this cannot be measured with the same precision, on account of the leakage of electricity, but even so, no departure from the law has been found with potentials as high as 200,000 volts. The wave-lengths corresponding to potentials of 4600 volts, 60,000 volts, and 200,000 volts are 270×10^{-18} cm., 20×10^{-8} cm. and $6{\cdot}1 \times 10^{-8}$ cm. respectively.

Line or Characteristic Spectra. Superimposed upon this continuous spectrum is another one consisting of a few very well-defined lines : their intensity is very large compared with that of the continuous spectrum. These lines depend upon the particular element used as anticathode. But if these lines, characteristic of each element, have different frequencies from one element to the next, their frequencies are nevertheless related to one another by a very simple law. Cathode rays are caused to impinge on anticathodes made from various elements belonging to the periodic system, and the wave-lengths of the characteristic lines determined in each case. **Barkla** had already noticed that, in each instance, there were two groups of lines very close to one another ; their frequencies increased as the atomic weight of the element investigated became larger. He called these the K and L radiations. Later on a third group known as the M group was discovered with elements of higher atomic weight ; later still the N, O and P groups were found, but these have only been observed with elements of very high atomic weight. In 1913 **Moseley** investigated the frequencies of the K and L series for all elements then available. The results he obtained have been most important in developing modern theory concerning the structure of atoms. We shall first give his results in their simplest form, afterwards returning to discuss them more fully.

When Moseley considered the most intense lines of the K series of all the elements, he found that their frequencies could always be represented by the formula

$$\nu = Rz^2\left[\frac{1}{1^2} - \frac{1}{2^2}\right] \quad . \quad . \quad . \quad . \quad (185)$$

The numerical value of R in this equation has the same value as in Balmer's formula for hydrogen. For z it was necessary to insert an integer the value of which was characteristic of the element in question. In fact, **this whole number is simply the number which the elements have in the periodic system, when hydrogen is denoted by 1, helium by 2, and so on, until uranium is reached with its number 92** [*cf.* Table 4, page 108].

The formula which holds for the frequencies of the strongest lines in the L series is

$$\nu = Rz^2\left[\frac{1}{2^2} - \frac{1}{3^2}\right] \quad . \quad . \quad . \quad . \quad (186)$$

But this means that the frequencies of X-rays can be represented by formulæ similar to those which Bohr used in connexion with the emission of light rays—in fact, the strongest line in the K series will be emitted when an electron returns from the second to the first orbit; the strongest line of the L series will be sent out when the electron jumps from the third into the second orbit. Similarly the M and N series arise when an electron passes into the third and fourth orbits respectively.

The way in which these rules were deduced from the experiments was somewhat as follows. If ν is the frequency for an element whose atomic number is z, whilst ν' applies to the next element $z + 1$, we have, from equation (184), if $\left[\frac{1}{1} - \frac{1}{2^2}\right]$ is written as $\frac{3}{4}$,

$$\nu = \frac{3R}{4}z^2, \text{ so that } \sqrt{\nu} = z\sqrt{\frac{3R}{4}}.$$

Similarly

$$\nu' = \frac{3R}{4}(z + 1)^2, \text{ or } \sqrt{\nu'} = (z + 1)\sqrt{\frac{3R}{4}}.$$

Hence

$$\sqrt{\nu'} - \sqrt{\nu} = \sqrt{\frac{3R}{4}}.$$

This equation means that when we pass from any one element to the next, the square root of the frequency of the corresponding X-ray line always increases by the same amount. If, therefore, a curve is drawn in which ordinates denote square roots of frequencies, whilst abscissæ denote atomic numbers, then a straight line is obtained. The slope of this line is different for each series of lines considered (Fig. 164).

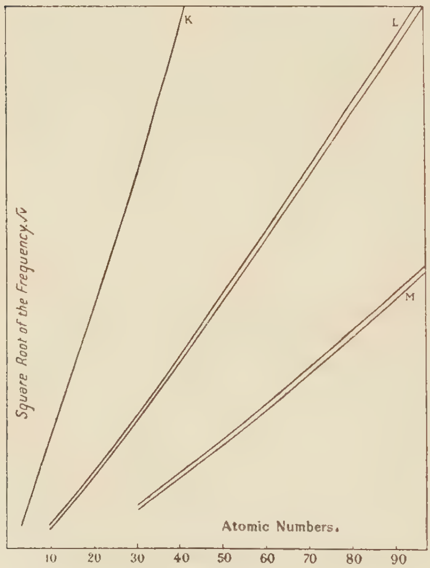


Fig. 164.

But the curves show more than this. The lines belonging to the K series, just like those of the L and M series, were not simple, but consisted of several lines whose frequencies differed from one another by an amount which was less than that between the lines of other series, but the intensity of the additional lines was very feeble in comparison with that of the chief lines. Bohr's theory was able to account for them all. If, in the general equation (180), n is put equal to 1, then by putting $m = 2, 3, 4$ . . . all the lines of the K series are obtained; $n = 2$, and then $m = 3, 4, 5,$ . . . gives the L series, $n = 3$ the M series, and $n = 4$ the N series.

When z is given values equal to the various atomic numbers, all the X-ray lines for every element in the periodic system are obtained.

The above remarks would apply strictly, only if the rotating electron were under the influence of the charge on the nucleus alone. Naturally this is not so, for the other electrons will also exert an effect, the sense of which will be such that the charge on the nucleus will appear to be less than what it is actually. It would therefore be more correct to write, instead of z, a quantity $(z - a)$. If we are justified in assuming that, for elements of high atomic number, the inner electrons are arranged in the same way, for each element, then we may assume that a is the same for each such element. Then, when differences between frequencies are considered, this term will disappear, so that the final result is as before. Experiment has confirmed this view.

Perhaps, on account of the important rôle played by these laws in framing an atomic theory, it is permissible to submit the subject in a somewhat different form. In so doing due attention will be paid to the 'screening action' which the electrons exert upon the nucleus.

Let us first of all consider only those radiations which are sent out when an electron jumps from the second to the first orbit for the various elements. It is therefore necessary for us to put $n = 1$ and $m = 2$ in Bohr's formula; so that $\left[\frac{1}{n^2} - \frac{1}{m^2}\right] = \frac{3}{4}$. The distinction between two elements is then only represented by a different value for Q [*cf.* equation (179)]. Q is no longer the charge on the nucleus, but that charge diminished by an amount due to the screening action of the electrons, *i.e.* the effective charge. We now make no assumption concerning the magnitude or constancy of this screening effect. If all the constants and known terms in the formula are collected together and we write in short

$$A = \frac{2\pi^2\varepsilon^2 m}{h^3} \cdot \frac{3}{4},$$

we obtain simply

$$\nu = AQ^2.$$

For another atom whose effective charge is Q_1 we have $\nu_1 = AQ_1^2$.

From these two equations we have

$$\sqrt{\nu/A} = Q \text{ and } \sqrt{\nu_1/A} = Q_1,$$

so that

$$\sqrt{\nu_1/A} - \sqrt{\nu/A} = Q_1 - Q.$$

Since A is known, and ν and ν_1 can be obtained from measurements made on the radiations emitted, we can determine the difference $Q_1 - Q$ for any two elements. *e.g.* any two consecutive elements of the periodic system. The results so obtained were very satisfactory. In the first place, the increase in charge from one element to the next is always the same, *i.e.* the increase in effective charge from one element to the next is always the same identical amount. In the second place, the magnitude of this increase can be calculated, and it is found to be very nearly

$$Q_1 - Q = 4{\cdot}77 \times 10^{-10} \, E.S.U.$$

The effective charge therefore increases from one element to the next by an amount numerically equal to that on an electron. From this we may conclude that, in all probability, the screening effect for two successive elements is approximately the same, and that this increase is due to the increase in the charge on the nucleus, which must change from one element to the next by an amount numerically equal to that carried by an electron. It is therefore possible to denote all elements by a number which specifies the number of positive charges on the nucleus.

Further, since for the atom of hydrogen Bohr's formula is in complete accord with Balmer's when the charge on the nucleus is equal to one unit positive charge, whilst the nuclear charge for a helium nucleus in the case of an alpha particle has already been proved to consist of two such charges, it follows that the charge of the nucleus of each element in the periodic classification is given by ascribing to the nucleus following one already known, a charge greater by unit amount. Then the nuclear charge number for all the elements is to be put equal to the atomic number, from hydrogen 1 up to uranium 92. For those elements which were still unknown, X-rays provided a means whereby, if a place had been omitted in the table, the fact that the next known element had an atomic number greater by 2 could be verified. If these same ideas are afterwards applied to the remaining lines in X-ray spectra, by putting for the next lines in the K series $n = 1, m = 3$, so that $\left[\frac{1}{n^2} - \frac{1}{m^2}\right] = \frac{8}{9}$, the result is just as before, *i.e* the difference in charge carried by the nuclei of two successive elements is always the same, in fact it is equal to the unit positive charge.

For the lines belonging to the L series one puts $n = 2$, and then $m = 3, 4, 5, \ldots$ Similarly for the M and N series. The result is always the same. Thus all the lines require the same uniform increase in the nuclear change from element to element as does a single line, and this is very convincing support for this method of representing the elements by successive numbers.

The Emission of Röntgen Rays. The lines of the K series are produced when a cathode ray ejects an electron from the innermost ring of electrons belonging to an atom in the anticathode, and its place is taken by an electron belonging to an outer ring. In general, this incoming electron will arrive from the second shell, so that this will be the most intense radiation in that particular series of lines ($m = 2$). But sometimes an electron will return from the third shell into the first ($m = 3$). When this happens the available energy will be greater, so that the frequency of this second line will be greater than that of the first ; but since this happens more rarely, the line is weaker. On still fewer occasions an electron will jump from the fourth orbit ($m = 4$), but when it does occur the frequency of the line will be greater.

In quite an analogous manner the L lines are produced when an electron is removed from the second orbit and its place taken by an electron from one of the 3, 4, . . . orbits. Similar remarks apply to the M and N series, each of which corresponds to the orbit into which the electrons return.

In Fig. 165 these jumps are represented graphically. The distances between the orbits are quite arbitrary.

It has thus been shown that equation (179), which does not apply to light waves from the heavier elements, is valid when used in connexion with X-rays for any element of the periodic system. This is to be explained by assuming that for those electrons which are near to the nucleus the effect of the nucleus on them is not appreciably disturbed by the more remote electrons. These remarks apply in particular to the heavier atoms, in which case the attraction by the nucleus is especially strong, so that the first orbit is particularly near to it. Nevertheless it is very remarkable that X-rays and light waves prove themselves to be essentially the same for light elements. For the smaller the atomic weight is, the longer the X-rays become; for the lightest elements, hydrogen, helium and lithium, the X-rays belonging to the *K* series lie in the ultra-violet region whilst some of those for the *L* series actually lie in the visible region of the spectrum, and, as such, had long been recognized. For hydrogen the lines are those of Balmer's formula. But these known lines, apart from the 'screening effect' due to other electrons, which is absent for the highest elements, arrange themselves in an ordinary X-ray series.

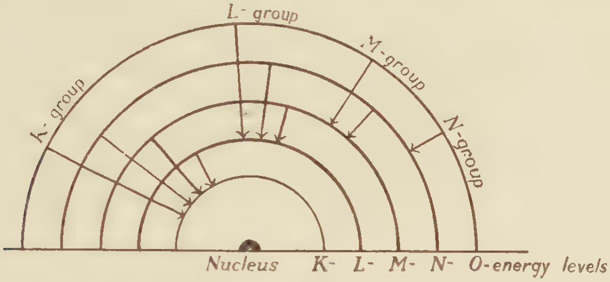


FIG. 165.

Secondary X-rays. We are able to produce light-rays by allowing light of the corresponding wave-length to fall on a body. When this happens resonance waves are obtained. X-rays behave in exactly the same way. If X-rays are allowed to fall on a body whose *K* or *L* radiation has the same frequency as the incident radiation, so that this body acts as a resonator for the oncoming waves, this body will absorb this incident radiation, and itself emit the so-called secondary radiation. The wave-length of this radiation agrees absolutely with that of the incident waves. Should this radiation fall on another body whose characteristic radiation is again the same as that which is incident upon it, then a third radiation is produced.

Röntgen rays are also capable of producing the so-called photoelectric effect, mentioned on page 295. The process is the reverse of that in which light is produced by cathode rays, for the incident light radiation causes an electron to be liberated from the metal. Even X-rays do this, and when it does happen, the electrons set free have imparted to them a definite velocity determined by the frequency of the incident X-rays according to the relation $h\nu = V\varepsilon$. This relationship is explained as follows. When X-rays of frequency ν fall on a metal, the energy quanta are equal to $h\nu$. If such a quantum is used to liberate an electron and then hurl it forth, a portion of the energy will be utilized *in separating* the electron from its atom. But since, for X-rays whose quanta are large, this work is so small that it can be neglected entirely, the whole energy will be available for accelerating the electron and can be measured. The method of doing this is the same as that used by **Millikan** in his experiments on the photoelectric effect. An opposing potential is applied which just prevents all the electrons from escaping from the plate which is being irradiated. The results obtained showed that the highest velocity acquired by electrons set free by means of X-rays is exactly equal to the velocity of the cathode rays which originally excited the X-rays. The fact that it is only the maximum velocity which has to be considered arises from considerations which have already been mentioned—all the subsidiary influences, such as the collisions which occur with other particles, or the retardation of the electrons in the layers of metal through which they pass, will diminish the velocity, but it can never be increased.

The equation $h\nu = V\varepsilon$ can be read just as well from right to left, as from left to right. In one case it governs the emission of X-rays by cathode rays; in the other it governs the excitation of cathode rays (secondary cathode rays) by X-radiation. Both phenomena obey the quantum laws, and both can be used to determine the magnitude of Planck's constant h. The values so obtained are concordant.

The confirmation of the quantum hypothesis by X-rays goes much further than can be discussed here. When the technique was improved whereby the spectral lines were observed, it was discovered that many of the earlier lines really consisted of several lines. These lay very close to one another, but they could all be measured with precision. When the ideas which Sommerfeld had used in his extension of Bohr's theory were applied to the fine structure of X-ray lines, an explanation of them was at once forthcoming. To-day we are able to consider this whole field as having been completely explored, since for each line in the X-ray spectrum of any element we can state definitely from which orbit an electron was sent out and which orbit it jumped into in order to excite this line. Tables have now been prepared in which all these orbits and the frequencies of the emitted waves are given.

CHAPTER IX

THE STRUCTURE OF THE ATOM

THE second main section in this book concluded with the remark that all bodies are built up from very small particles which we call atoms. In the third section we have made a synopsis of all we now know concerning the atoms themselves. In our attempts to unravel the mystery of the atom we have been aided by observations of the most varied phenomena in physics.

First of all we found that there are charges of negative and of positive electricity in all bodies. The negative electricity has been shown to consist of very small charges which never appear in a subdivided form. Whether they occur in electrolysis, in cathode rays, in the beta rays from radioactive bodies, in gaseous ions, or as free negative charges liberated by the action of ultra-violet light on a metal, they are always found to carry the same charge, and everywhere their nature is invariable.

But since atoms as a whole are uncharged, it follows that there must be a positive charge inside them numerically equal to that of the extranuclear electrons.

The positive and negative charges in atoms are separated from one another. Anyone who wishes to hold the opposite point of view would maintain that the electrons spread themselves out or at least possess continuity and can mix with the positive charges. Such an assumption would be quite arbitrary, and there is no single known fact which supports it. On the other hand, however, there are numerous phenomena which quite definitely contradict it. On many occasions we have succeeded in separating the two types of electricity from one another—in fact we may say that all phenomena which have led to the discovery of the electron, and which have just been mentioned, support this view. In addition we may mention the facts of electrostatic induction, which show that, under the influence of an external electric field, the two electricities move in opposite directions. In electrolysis we even see the two charges appearing spontaneously and moving about as ions.

All these facts require that the two opposite kinds of electricity shall be separated from one another by a finite distance even in an atom, since if it were otherwise they could not be separated from one another again by any finite force.

Since positive electricity has never been observed except in association with inert mass, we have no reason to suppose that in the positively charged portion of an atom one would be able to separate the positive electricity from the inert mass there present. We do not even know whether, in general, positive electricity is, of itself, an entity. All phenomena point to the fact that it is sufficient to assume that a body acquires a positive charge when it loses a negative one, just as it is not necessary for us to suppose that a vacuum is some new form of matter, for a vacuum arises whenever the air is removed from a vessel. We are therefore justified in stating that an atom consists of two parts—electrons and inertial mass. This latter constitutes the nucleus of the atom.

1. ELECTRONS

On account of the attraction which exists between negatively charged electrons and the positively charged nucleus, these two portions can only remain permanently separated from one another if some other force is present inside the atom which opposes the force of attraction. The centrifugal force, which must be present if we assume that the electrons revolve around the nucleus, is sufficient for this purpose. We have no direct evidence to show that the electrons do revolve, except that the previously stipulated force of attraction necessitates this motion. Moreover, any other repulsive force, which would have to be of the same magnitude as the force of attraction, would necessarily be capable of distinction in many ways, just like the attraction. Such another force is most certainly not present.

So far this is Rutherford's idea of an atom. According to his theory an atom is only a planetary system in miniature, in which the planets (electrons) revolve around the central sun (nucleus).

Let us examine this analogy more closely for a moment. Enormous spaces are available, in which the planets of the solar system may move. Only a very minute fraction of the space around our sun is occupied by the planets themselves. How does this apply to an atom ? Let us take as an example an atom of hydrogen. We know that the nucleus of such an atom has at most a radius of about 10^{-12} mm. The radius of the first electron orbit is $0{\cdot}5 \times 10^{-7}$ mm. In order for us to realize more easily what these numbers mean, let us multiply them by 10^{12}. On this scale the hydrogen nucleus has a diameter of 2 mm., the size of a pin's head. The nearest orbit would then be at a distance of about 50 metres. The whole space in between is a perfect void. If the scale is enlarged so that the nucleus is represented by our sun, the radius of which is about 700,000 km., then the nearest approach made by the electron would occur at a distance of $35{,}000 \times 10^6$ km. The remotest planet of the solar system is Neptune, whose distance away is less than 4500×10^6 km. This is only one-eighth part of the distance of the nearest electron (on the above scale). The space which the elements ' take up ' is therefore still less filled than the space in which the planets of the solar system move.

The forces which keep the whole together are the electrical forces, which, like those due to gravitation, decrease according to the square of the distance. Hence single electrons with their negative charges can only be removed from an atom by powerful external forces. When this occurs we no longer have an atom but a positively charged ion. Likewise one or more electrons can be added to the number usually present—we then have a negative ion. Multivalent ions, which the chemists had long recognized as those which could enter into combination with several monovalent atoms, were proved, in electrolysis, to be those ions which carried 2, 3, . . . electrons.

The charge on the nucleus is constant and a characteristic of each element. When helium, the second element, had been shown to possess two electrons (which are lost when the atom exists as an α-particle), and **Moseley** had succeeded in demonstrating that the X-ray spectra of all the elements could be accounted for by assuming that the charge on the nucleus increased by unity from one element to the next, *i.e.* the nuclear charge was equal to the atomic number, many striking deductions followed.

The Atomic Numbers. The discovery of these numbers enabled a more satisfactory classification of the elements to be made, for it was then recognized that the system was improved if the elements were arranged, not according to their atomic weights, but their atomic numbers. For whereas only an imperfect classification could be made in the first case, yet when the elements were arranged according to the simple ordinal numbers the whole system became one of great regularity. The new system justified itself at once. In those places where anomalies existed in the old system, as, for example, when the chemical properties of an element required that it should be placed in a position different from that which its atomic weight would indicate, yet, when the elements were arranged according to the frequencies of their X-ray spectra, these discrepancies disappeared and the arrangement was in complete accord with chemical data. It was only for six places that it was necessary to suppose that an unknown element was lacking. The justification for this assumption has since been forthcoming since X-ray spectra have been the means whereby some of these elements have been discovered. They were the elements, hafnium 72, masurium 42, rhenium 75, and illinium 61.

These discoveries were made possible by the laws which had been formulated in connexion with X-ray spectra. These laws have also played an important rôle in the development of theoretical physics. X-ray characteristic spectra are independent of whether the anticathode element exists alone or in chemical combination with other elements. Since the wave-lengths of these lines can be measured very accurately by comparing them with those of a neighbouring element, the presence of an otherwise unknown element can be established with certainty even if only a trace of it is available. After this has been done larger quantities can be obtained and easily identified by the usual chemical tests.

For our knowledge concerning matter we owe much to these characteristic X-ray spectra of the various elements, since in them we have a proof **that atoms continue to exist even when they enter into chemical combination.** The weight of this evidence overwhelms all that which comes from other sources. For until recently we had only been able to make conclusions concerning this problem from our knowledge of the conservation of mass and other general properties of matter. The whole series of various X-ray spectra, which we are able to measure more exactly than any other general property of matter, retain their individual characteristics, and this remains valid throughout all the elements, the entire system of the material world. For the *K* series of lines, which are produced by happenings nearest to the nucleus, this constancy holds with great exactness; for the *L*, *M*, and other groups, small but nevertheless quite clear and definite changes in wave-length are produced by the presence of neighbouring atoms.

The Quantum Theory and the Periodic System of the Elements. The study of atomic structure was approached by **Niels Bohr** by considering the light waves emitted by the various elements after he had established Balmer's formula in connexion with the light waves from a hydrogen atom. In the numerous conditions which can be produced in an atom by the outer electrons which move very readily, he realized the possibility of obtaining an explanation of the many light waves which an atom can emit. Since reasons were to hand why an atom of hydrogen may be assumed to possess one electron, on that account alone the spectrum of hydrogen rose in importance, because the complete motion of an electron around a nucleus had already been solved. We refer to the problem of two bodies, or the motion of two bodies which attract one another. The perfect agreement between observation and calculation at once placed Bohr's hypothesis beyond all doubt. For the remaining elements the calculation could not be made with the same degree of perfection, but a whole series of individual phenomena, which were confirmed by experiment, resulted in so many new conclusions, that eventually the arrangement and movement of the electrons of all the elements, even those with a large number of electrons, were successfully elucidated on the bases of these fundamental ideas. Even though details in this scheme will require to be modified and extended, our picture of the atom has been drawn with such careful attention to facts, that it is difficult to think that the theory will ever be replaced by another which is entirely different.

In arranging the elements, the inert gases were selected as a starting point. These gases never enter into chemical combination with other elements, nor do they under ordinary circumstances unite together to form molecules containing two or more atoms. From this fact, **Kossel** and **Lewis** simultaneously concluded that the extra-nuclear electrons of the inert gases must be held firmly together **in some closed configuration.** This assumption received support from the fact that the potentials necessary to ionize them are high. The elements which follow them in the periodic table [*cf.* p. 111], viz. *Na*, *K*, *Rb*, *Cs*, are monovalent electro-positive elements, *i.e.* they very easily part with one, but only one, electron. It was therefore concluded that this one electron is placed somewhat like an outpost beyond the remaining part of the atom, and that it is less rigidly bound than the other electrons. The inert gases are preceded by the electronegative elements *F*, *Cl*, *Br*, *I*, all of which are monovalent, *i.e.* they will all readily take up another electron in order to complete the electronic configuration characteristic of the inert gases. Two places after the inert gases are found elements which are electropositive but divalent; they easily part with two electrons. The elements which occur three places after the inert gases are trivalent. Similarly the elements which occur two and three places before the inert gases are respectively divalent and trivalent, but electronegative.

All these phenomena support the idea that the electrons are arranged in rings or shells around the nucleus. In each shell, however, there are only positions for a limited number of electrons; more electrons than these cannot be present, since the repelling forces between the electrons would then be too great for stability.

The shells of the inert gases are completely filled, whilst for the elements which follow them the outermost shell only has one electron. Since it is only lightly bound to the remainder of the atom, it can easily be separated from it. For the elements immediately preceding the inert gases the outer shells are deficient by one electron. The fact that there is a periodic return of similar properties amongst the elements is therefore due to the commencement of a new shell every time one has been completely filled.

The electronic shells nearest to the nucleus are the smallest, so that only a few electrons can be contained therein, whereas the others become larger step by step, thereby being able to offer more positions into which electrons can be placed; this is why the number of elements in the lower horizontal rows of the periodic table becomes more and more. The radii of these shells depends upon the attractive force between the nucleus and the electrons in them, so that for the heavier atoms, with their larger nuclear charge, these shells are smaller than for the lighter atoms. This also provides the reason why the heavier atoms, in spite of their greater mass and larger number of electrons, do not occupy a volume which is essentially greater than that for the lighter elements.

By paying very careful attention to all the facts concerning the physical and chemical behaviour of all the elements, **Bohr** has endeavoured to account for the number of electrons in the shells of the various elements. An epitome of his work is given in Table 34 on the following page. It does not follow, however, that the structure of the various atoms is known with absolute certainty, but the fact that such an attempt was ever made proves that the clues which have been obtained indicate that an atom is constructed according to a definite plan. For example, from these considerations, Bohr was able to predict the properties of the element hafnium, which was soon afterwards discovered.

Chemical Processes. From the remarks which have already been made, it follows that ions are produced in electrolytes when electrons are either removed from or added to the normal atom. Since the idea of valency goes together with the capacity which an element has for forming ions, we have to conclude that chemical actions are the result of changes taking place in the outermost electronic shell. This is also in accord with the fact that some chemical actions take place so readily, that some chemical compounds can be made by merely mixing the elements together. The electrons which are present in the incomplete outer shell of an atom are therefore known as valency electrons.

The emission of light is, on the whole, due to the motion of these outermost electrons. Light waves are therefore not determined by the atomic weight, *i.e.* the mass of the nucleus, but only by the charge on the nucleus and the number and arrangement of the outer electrons. We now see why it is that the isotopes of an element all give the same spectrum, although the masses of their nuclei are different.

In general it cannot be expected that the spectral lines for the heavier elements can be calculated from Bohr's formula, for the formula only considers the action between the nucleus and a single electron. But the motion of the outermost electron will depend to a large extent on the other electrons, the effect of which, we can suppose, will be to reduce the attraction of the nucleus, since they tend to screen the outermost electron from the nucleus. If the outermost electron happens to be in a very remote orbit we may assume that the other electrons and nucleus are all concentrated in one point at the centre of the atom. The total effect will then be as if one positive charge remained. The attraction on the remote electron will therefore be the same as if it belonged to the nucleus of a hydrogen atom. A glance at Table 34 will show that these remarks apply especially to the alkali metals *Na*, *K*, *Rb*, *Cs*. All of them have outside the shells which have been completed another shell, in which there is only one electron present. Since it is only this electron which is responsible for the emission of light, the spectra of these metals will only contain a few lines and, as a rule, they will be similar to that of hydrogen. They will, however, deviate from it in so far as the electron belonging to the elements of higher atomic weight can get amongst the nearer electrons and there experience a different attraction, since the other electrons will no longer be screening it from the nucleus.

TABLE 34

The Electron Orbits for some Elements

	1_1	2_1	2_2	3_1	3_2	3_3	4_1	4_2	4_3	4_4	5_1	5_2	5_3	5_4	5_5	6_1	6_2	6_3	6_4	6_5	6_6	7_1	7_2
1 H	1																						
2 He	2																						
3 Li	2	1																					
4 Be	2	2																					
5 B	2	2	1																				
10 Ne	2	4	4																				
11 Na	2	4	4	1																			
12 Mg	2	4	4	2																			
13 Al	2	4	4	2	1																		
18 A	2	4	4	4	4																		
19 K	2	4	4	4	4		1																
20 Ca	2	4	4	4	4		2																
21 Sc	2	4	4	4	4	1	2																
22 Ti	2	4	4	4	4	2	2																
29 Cu	2	4	4	6	6	6	1																
30 Zn	2	4	4	6	6	6	2																
31 Ga	2	4	4	6	6	6	2	1															
36 Cr	2	4	4	6	6	6	4	4															
37 Rb	2	4	4	6	6	6	4	4			1												
38 Sr	2	4	4	6	6	6	4	4			2												
39 Y	2	4	4	6	6	6	4	4	1		2												
40 Zr	2	4	4	6	6	6	4	4	2		2												
47 Ag	2	4	4	6	6	6	6	6	6		1												
48 Cd	2	4	4	6	6	6	6	6	6		2												
49 In	2	4	4	6	6	6	6	6	6		2	1											
54 X	2	4	4	6	6	6	6	6	6		4	4											
55 Cs	2	4	4	6	6	6	6	6	6		4	4				1							
87	2	4	4	6	6	6	8	8	8	8	6	6	6			4	4					1	
88 Ra	2	4	4	6	6	6	8	8	8	8	6	6	6			4	4					2	
89 Ac	2	4	4	6	6	6	8	8	8	8	6	6	6			4	4	1				2	
90 Th	2	4	4	6	6	6	8	8	8	8	6	6	6			4	4	2				2	
91	2	4	4	6	6	6	8	8	8	8	6	6	6			4	4	3				2	
92 U	2	4	4	6	6	6	8	8	8	8	6	6	6			4	1	4				2	

Spark Spectra. By means of electric sparks it is very easy to remove one or even several electrons from the region in which they are attracted by the nucleus, *i.e.* the elements become ionized. When the now outermost electron is excited so that light is emitted, the conditions will be very different from those actuating in the uncharged atom. If, for example, an atom of helium is ionized, there is only one electron left which revolves round the nucleus as in a hydrogen atom. In this instance Bohr's simple deduction of the hydrogen spectrum will apply to that of the *He*-ion, if the charge on the nucleus is doubled. We have already learnt that these ideas are confirmed by experiment [*cf.* p. 349].

The atoms of the alkali metals are also very easy to ionize since they readily surrender their outermost electron. When this occurs a glance at Table 34 shows that the electronic configuration is the same as that in the corresponding inert gas. Sodium becomes just like neon, potassium like argon, rubidium like krypton, and cæsium like xenon, in so far as the electronic configuration is concerned. All the inert gases (except helium) have a spectrum which is very rich in lines on account of the great number of electrons present in their outermost shell. When, therefore, the alkali metals are excited by means of a spark discharge, in place of spectra similar to that of hydrogen and poor in lines, there are suddenly obtained spectra which are very rich in lines and which have the characteristics of spectra belonging to the inert gases.

When the so-called metals of the alkaline earths (divalent), *Mg* 12, *Ca* 20, *Sr* 38, *Ba* 56, are ionized, they have the same electronic arrangement as pertains to the alkali metals; their spectra are therefore not rich in lines.

Millikan and **Bowen** have succeeded in producing ions of doubly ionized boron (atomic number 5), *i.e.* two of the three electrons normally present in the outer shell were removed. The spectrum then became the same as that of lithium, which is similar to that of hydrogen.

Fowler eventually succeeded in producing triply ionized silicon (14), whereupon its spectrum became similar to that of sodium (11).

Since in all these experiments we are actually dealing with ionized atoms, this fact could be demonstrated, for example, by showing that when such atoms were excited they were deviated by electric forces in the case of cathode rays.

X-rays. For those electrons in the innermost shell, the screening effect, to which we have already referred, is either entirely absent or else so reduced that it is negligible. It is for this reason that the emission of X-rays, which is due in particular to electrons in the first shell, can be represented in such a simple manner for all the elements. Even the X-ray spectra of isotopes are identical, since the mass of the nucleus exerts no action on them.

Perhaps in the *L* and *M* lines in X-ray spectra the other electrons will exert a small influence on the radiations. But for all those elements in which the shells in question are completely occupied, this influence will be of practically the same amount. Hence, in the approximate treatment here presented, this effect has been disregarded.

Magnetism. We are now led to the idea that the atomic currents, whose presence was necessary in order for us to develop a consistent theory of magnetism, are to be found in the revolving electrons themselves. Circuits through which a current is flowing are equivalent to a magnet whose axis is perpendicular to the plane of the circuit [*cf.* p. 257]. If therefore an atom is exposed to the action of a magnet (*i.e.* placed in a magnetic field), the small atomic magnets and the external field will exert an influence on one another; it is therefore fairly obvious that all known phenomena in connexion with magnetism will depend upon this reciprocal action between the electron orbits and the external field.

Paramagnetism. Before Rutherford formulated his atomic model, **Langevin**, in 1905, assumed that positive and negative charges rotated in circular orbits in the atoms. From this assumption he had drawn certain conclusions, which later on became very important. If we now call his charges electrons, his line of thought was somewhat as follows.

The orbits of the various electrons, in general, lie in different planes, each of which will therefore have a different magnetic axis. An external field will therefore rotate these manifold axes so that they all tend to lie in the direction of the applied field. If we now assume that the electron orbits of an atom have fixed positions relative to one another, then every turning moment exerted upon an electron orbit must act on the whole atom. But the various forces tending to produce rotation will very often oppose one another, and therefore disturb the effects which they would otherwise produce. If, for example, the electron orbits were arranged quite regularly around the nucleus, then the directional forces which would be exerted on them would counterbalance one another, and the atom would apparently not respond to magnetic influences : it would be non-magnetic. But if the orbits are not distributed absolutely regularly around the nucleus, then these forces will not counterbalance one another completely, so that the atom will acquire a resultant magnetic moment, the axis of which we must call the magnetic axis of the atom. Under the action of an external field, all such atoms will be rotated, so that this direction tends to coincide with that of the applied field (just as a freely suspended needle rotates into such a position that its axis points in the direction of the earth's field). Such atoms will belong to paramagnetic bodies.

When all the atoms have been rotated so that their axes all lie in this direction, we have the phenomenon of magnetic saturation. But this condition will be opposed by the thermal movements present in a body, and since these irregular motions become more pronounced as the temperature rises, it follows that the maximum amount of magnetism which a body can acquire will decrease as the temperature increases. This is in accord with the behaviour of all paramagnetic substances (the first law of **Curie**).

Weiss and **Gans** have been able to perfect the connexion between these ideas and observed facts. They followed **Ewing's** example [*cf.* p. 231], and took account of the mutual actions between neighbouring poles, whereby Gans, basing his remarks on certain assumptions, even denied the existence of a magnetic fluid ; he considered that these mutual effects between electric charges in motion and the external field are sufficient to explain magnetic phenomena. Quite recently (1927) **Honda,** of Tokyo, has proposed the following theory to account for the presence of paramagnetism. He assumes that electrons which are embedded in the nucleus [*cf.* p. 378] rotate so rapidly that a sufficiently large magnetic moment is produced.

When Weiss determined the degree of magnetic saturation at various temperatures, he not only found the variations which were expected, but he also thought that he had noticed that the change did not take place regularly with temperature : it took place in jerks, similar to the way in which Millikan had found that the velocity of his oil drops changed when they received an electric charge. From this he concluded that there was a limit to the smallness of a magnetic moment. He attributed it to the individual atom, and called it the '**magneton**'.

In Bohr's theory of the atom, where the charges, radii of the orbits, and the velocities of the electrons are all fixed, it is possible to calculate as a matter of course the magnetic moments due to the individual electrons. It is then found that the magnetons are approximately fixed in size, and that an integral number of them must always be present in the different atoms. The magneton according to Bohr's theory is about five times as large as that assumed by Weiss. Finally, in 1925, **Uhlenbeck** and **Goudsmit** used the idea which **Compton** had already suggested, that the electrons rotate about their own axes, in order to explain certain effects associated with the fine structure of spectral lines, and they were also able to explain a large number of other details concerning spectra. But these rotating electrons would give rise to very small elementary magnets at the same time, and therefore have to be considered in any theory purporting to explain magnetic phenomena.

Diamagnetic bodies are characterized chiefly by the repelling forces which exist between those bodies and the external field which excites them. Magnetic actions between two diamagnetic bodies alone, without an external field, have never yet been observed. One might assume that diamagnetism is first produced at the moment when the external field begins to act upon the body. If so, diamagnetism would have to be counted amongst phenomena due to induction. Such induced phenomena must indeed arise when the electronic orbits are brought into a magnetic field. The direction of the current which is induced will be such that it opposes the exciting field. According to the tenets of the theory of electrons, *e.g.* the views of **Lorentz** which were referred to on page 268, a current will, in general, never be actually produced by induction, it will only be directed by it ; the current must be present already, exactly as in the case of electrostatic induction, where the charges are not produced but only redistributed. If such electron currents are brought into a magnetic field, then according to the laws of electromagnetic induction [*cf.* p. 264] the currents will be so directed that repulsion takes place between them and the field, exactly as the phenomena of diamagnetism require. This action will take effect on all the electronic orbits irrespective of their orientation, and will therefore be independent of any random motion of the atom as a whole. It therefore follows that diamagnetic phenomena will be independent of temperature,. and in general, all substances must be expected to be diamagnetic. These diamagnetic effects will be counterbalanced by any paramagnetism present, so that a body will appear to be diamagnetic or paramagnetic according to whichever of these two effects predominates.

In metals we have to assume that, in addition to the revolving electrons, there are some which are free and which can move quite freely along a rectilinear path for some distance. Under the action of a magnetic field this path will become curved [just like the path of an electron when it enters a magnetic field (*cf.* p. 299)]. The total effect of these curved circuits will be to impart to the metal a magnetic moment whose direction is opposite to that of the applied field. Attempts which have been made to calculate the intensity of this diamagnetic effect have so far been unsuccessful.

Modern Views concerning the Arrangement of Electrons. The above-mentioned results, in combination with others, have enabled a considerable number of new laws to be formulated in recent years : we shall only mention the most important of them. For whereas two quantum numbers have hitherto been used to denote the motion of an electron in a plane, a third quantum number now manifests itself. It concerns the position of the complete electron orbit and therefore its inclination with respect to a magnetic axis. A fourth quantum number refers to the revolution of the electron which can occur in two directions.

These ideas receive much support from the experiment of **Stern** and **Gerlach** in which silver atoms were used [*cf.* p. 203]. A narrow beam of vaporizing silver atoms gave, when no magnetic field was present, a distinct linear image of a slit through which they passed. The silver atoms were then allowed to pass by the edge of a strong magnet. It was then to be expected that the individual atoms in the beam would either be repelled or attracted by the magnet, according as the outermost orbit in which an electron is revolving presents a north or south pole to the strongest part of the magnetic field. The experiment showed actually that the beam was divided into two portions, of which one was attracted by the magnet, the other being repelled.

Pauli's Exclusion Principle. For determining the distribution of electrons in the various quantum shells according to these fourfold quantum numbers, the rule due to Pauli has many times proved itself of value. It states that of all the electrons belonging to an atom, two or more electrons can never have attributed to them the same four quantum numbers. This does not contradict the statements which have already been made, according to which, for example, Na has 2×4 electrons in the second shell. According to the above rule, the four electrons must differ from one another only in the third and fourth quantum numbers.

In 1926 **Fermi** extended this rule so that it could be applied to large bodies. According to him, in any closed space containing gas all the molecules must possess a different quantum content. This rule should have a considerable influence on any theory concerning the behaviour of gases, especially at very low temperatures. Even at the absolute zero all the molecules could not lose their entire energy but only one of them. If N denotes the number of molecules in the gas space, then at least one molecule must possess N energy quanta, another $N - 1$, etc. Accordingly, at very low temperatures, gases would no longer obey the laws to which we have previously supposed them to conform ; the specific heat would become extremely small as required by Nernst's law. Gases are then said to be 'denatured'.

The above rule is one which has actually proved itself of great service, but an explanation of the rule, say from the properties of electrons, has not yet been found.

This question is being tackled in all countries with great zeal, so that still further important progress may be expected from these combined efforts. **Ultimately all this work should result in deriving all the physical and chemical properties of the elements from the number and arrangement of the electrons around the nucleus. That such an explanation, based on certain definite principles, is possible, is to-day no longer doubted.**

When the first attempts were made to arrange the elements in a system, it was their atomic weights which were the important factor: to-day it is the charges on atomic nuclei. But since, except for a few small differences, the order of the elements was the same in either system, it is to be concluded that there is some intimate connexion between the atomic weight and the nuclear charge. We shall deal with this relationship in the sections which follow.

2. THE NUCLEI OF ATOMS

As we have had occasion to reiterate several times, it is mainly the nucleus of an atom which is responsible for its mass, and hence, ultimately, for the whole mass of the body. Even when the electrons number about one hundred, as in the heavier atoms, their contribution to the mass of the atom is negligible. For even one hundred electrons have a mass which is only about one-twentieth that of a hydrogen atom, so that for an element whose atomic weight is 200 this same number of electrons would only contribute 0·025 per cent. of its mass. For hydrogen, of atomic weight unity, the mass of its one electron is only $\frac{1}{1800}$, or approximately 0·05 per cent. The error which accompanies the experimental determination of atomic weights is less than this only in very rare instances.

What else can we assert concerning the nucleus of an atom ? Physics is only able to state just as much as has been observed. In the early days of modern science, it was thought that the atoms or molecules as a whole were responsible for chemical actions ; this view has now been entirely discarded. It is now believed that it is to the electrons belonging to the outermost shell that chemical actions, as well as the emission of light, must be attributed, whilst those electrons which are nearest the nucleus are responsible for the production of X-rays. In all these processes the nucleus is only concerned in so far as it determines, through its electrical charge, the number of extra-nuclear electrons and their orbits. But it contributes nothing to the problem as to whether or not it is possible for us to distinguish any further entities in the nucleus.

Generally speaking, there are three phenomena for which the properties of the nucleus are directly responsible ; we refer to radioactivity, the regularities which occur in connexion with atomic weights, and finally to the artificial disintegration of atoms by alpha particles. Moreover, the phenomena due to the passage of alpha and beta rays through matter have provided us with some knowledge concerning atomic nuclei.

The phenomena of radioactivity are due to changes taking place inside the nucleus of the atom. This is at once apparent, for the emission of alpha particles having an atomic weight equal to four can only take place from the nucleus itself. We have already dealt with this in the section on radioactivity. From the fact that it is only the nuclei of helium atoms which are sent out by all the twenty-three radioactive bodies known to emit alpha rays, we conclude that everywhere in the parent element from which they originate the constituents of the alpha particles must be held more firmly together than to the remaining parts of the atom. This implies that we have to consider *He*-nuclei as being structural elements of the material world, not only for radioactive bodies, but for all elements, on account of the reasons already adduced [*cf.* p. 322].

The Laws of Radioactive Changes. When an alpha particle, which always carries with it two charges of positive electricity, is emitted by an atom, that atom no longer contains positive and negative electricity in equal amounts. But since the damage to the inaccessible nucleus cannot be repaired, equilibrium is restored in the electron shells by the removal of two extra-nuclear electrons from them. This means that we have a new element whose nuclear charge and number of outside electrons are each less by two : the new element will therefore have an atomic number which is lower than that of the parent element by two. In fact, the new element is found to belong to that position in the periodic system which its whole chemical behaviour would indicate. This is very clearly shown in the case of the decay of radium into radium emanation which is, as it ought to be, an inert gas. According to the rules previously given, it should fall together with the other inert gases in a vertical column of the classification.

Moreover, we have to assume that, in addition to the *He* atoms in the nucleus, there are also electrons present, for the beta rays also can only have their origin in the nucleus. It is also true that the separation of an electron from the outer shells of electrons can take place ; this, however, does not produce a new element, but only an ion of the element in question. As soon as the electron is replaced the original atom, with all its properties, again appears. When a β-particle is ejected, however, something quite new is produced. The old atom never remains ; instead, we have a new element with a nuclear charge which is greater by unity than for the parent element. Actually the charge on the nucleus has diminished by one electron, but this is equivalent to an increase by unity of its positive charge. The new element attracts an electron to its outer shells from somewhere in its vicinity, so that the whole is again in electrical equilibrium. But at the same time the nuclear charge and number of electrons have each increased by unity, so that the atomic number of the new element is one higher than that of the original substance.

The emission of alpha and beta particles therefore produces a regular rearrangement of the elements. If the emission of an alpha particle, which often takes place from an element belonging to the radioactive series, is followed by the emission of two β-particles, then by the first process the atomic number of the new element will be reduced by 2 ; the second process will raise this number by unity at each stage. Thus the final element will have the same place in the periodic system as the first element ; it will have the same nuclear charge and number of extra-nuclear electrons, but its mass will be less by an amount equal to that of an alpha particle. The new element is an **isotope** of the first element, with an atomic weight differing from it by four units, but with the same charge on the nucleus. These two elements will exhibit the same chemical properties, and be characterized by both light and X-ray spectra which are identical in each instance. In a similar way other elements occur amongst the radioactive substances, which consist of groups of isotopes ; there may be as many as nine for the same place in the periodic system. These laws concerning radioactive decay were simultaneously proposed in 1913 by **Soddy** and by **Fajans.**

If, therefore, a change of four units in atomic weight leads to an element only two places removed from it, it would be expressed that, in general, the change in atomic weight from one element to the next would be two units ; and since helium itself has an atomic number 2 and an atomic weight 4, it follows that every alternate higher element should have a nucleus which is richer by one helium nucleus. The atoms whose atomic numbers are even, therefore, may consist entirely of helium nuclei. On the other hand, when this number is odd, there will be present in the nucleus at least one or a few hydrogen nuclei which have not combined together to form a helium nucleus.

The atomic weight of the helium nucleus is, curiously enough, four times that of hydrogen. But the nucleus of a hydrogen atom possesses always only a single positive charge. If we therefore assume that the helium nucleus consists of hydrogen nuclei, it would have 4 positive charges on its nucleus ; since it only has 2, the other two positive charges must have been neutralized by two electrons embedded in the nucleus. This is all the more probable since the *He* nucleus, if it did consist only of hydrogen nuclei (which would repel one another), could not have a permanent existence. But we have already observed that alpha particles are held very rigidly together. We therefore have to assume that there are two electrons present with four hydrogen nuclei in the nucleus of helium, and that they form, as it were, the binding material. Heavier atoms which consist of a certain number, n helium nuclei, will have an atomic weight 4 n. The corresponding positive nuclear charge 4 n will be partly neutralized by 2 n electrons present in the nucleus, whilst 2 n extra-nuclear electrons will render the atom neutral. In general, if M is the mass of an atom's nucleus, then it consists of M hydrogen atoms with M positive charges. If A is the atomic number for the element in question, then its atom will have A electrons revolving around its nucleus, the remaining $M-A$ positive charges in the nucleus being neutralized by the same number of electrons which must be present there. These electrons also enable the different hydrogen nuclei in the nucleus to remain together in such a small space.

The nucleus of a hydrogen atom, which is the smallest electropositive element of construction, is called a **proton,** but the manner in which these protons are arranged with the nuclear electrons is unknown to us.

The idea that atomic nuclei consist of hydrogen nuclei has received much support from two remarkable sources in recent years. Both have already been discussed. One is the discovery of the fact that those atomic weights which are fractional are really the mean values of two or more components, each of which has an atomic weight which is an integral multiple of that of hydrogen [*cf.* the discussion on isotopes, pp. 215 *et seq.*]. The other concerns the artificial disintegration of atomic nuclei by means of alpha particles, during the course of which hydrogen nuclei are released from the various known elements.

It is quite certain that the nuclei of all atoms do not consist entirely of helium nuclei, for otherwise their atomic weights would all be of the form 4 n. It is therefore to be assumed that, at least for all elements whose atomic number is odd, their nuclei have present in them hydrogen as well as helium nuclei. For example, nitrogen, atomic weight 14, atomic number 7, must contain besides three helium nuclei, two hydrogen nuclei. The 14 positive electric charges will require 14 electrons in order for the whole atom to be neutral ; since only 7 electrons lie outside the nucleus, 7 others must be situated in the nucleus itself. Of these three times two can be embedded in the three helium nuclei, but the seventh will be lodged elsewhere in the nucleus.

It appears, however, that the single hydrogen nuclei are not so rigidly constrained to the rest of the nucleus as those which have combined together to form a helium nucleus. At least the disintegration of elements whose atomic weights are a multiple of 4 is only obtained with greater difficulty than for the others whose atomic numbers are odd.

Hitherto the heavier elements have not been disintegrated artificially, a fact which does not justify us in assuming that they are more rigidly built than the lighter atoms. A much more legitimate assumption has been made by **Sommerfeld**. According to him the heavier atoms, on account of their strong positive charges, repel the alpha particle which is hurled at them, with such force that its whole velocity is annulled before a collision with the nucleus can take place. In the case of lead, for example, the alpha particle cannot approach the nucleus more closely than $2{\cdot}9 \times 10^{-12}$ cm. This is the same distance as that deduced from the passage of alpha rays through matter.

The idea that the protons in a helium nucleus are held together in an especially rigid manner finds additional confirmation from a fact which comes from quite another source. According to the views which have already been discussed concerning the formation of the whole atomic system, those elements will be most abundant which have the longest existence. If the elements to which we have resource are arranged according to the frequency with which they occur, then one-half of all matter is oxygen ($O = 16$), which can be conceived as being built up of 4 *He* nuclei. According to the frequency with which it occurs, we might say that oxygen is the most rigidly constructed of all the elements.

Further, ninety-nine per cent. of all available matter consists of elements whose atomic weights are less than that of iron (56), those with even atomic numbers occurring more frequently than the others. This would indicate that elements whose nuclei consist of helium nuclei alone occur more often than do those with hydrogen nuclei as well as helium nuclei. By extending these investigations to meteorites, which are very important in this connexion, since they have been less often in contact with other matter, **Harkins** obtained the result that 97·6 per cent. of all elements consist of atoms having even atomic numbers. The elements of the periodic system which have not yet been discovered all have odd atomic numbers.

All these facts point, without exception, to the hypothesis that the nuclei of all elements are made up of hydrogen nuclei and electrons.

3. Atoms and Molecules

Our knowledge of the elements will be considerably extended if we can find out the manner in which atoms unite together to form molecules. Within the last decade some not unimportant progress has been made in this direction. It is only in very exceptional circumstances that we find elements existing in their natural state as atoms ; generally, several of them have united together to form molecules.

1. When modern science was still young it was believed that the atoms, as such, had disappeared when a molecule was formed, and that an essentially new body had been produced. The grounds on which later investigators have asserted that atoms persist as atoms in molecules were as follows. In the first place such an assumption is very simple ; and it is a general principle that we are not entitled to assume a fundamental change if a phenomenon can be explained without assuming any such change. This conception arose out of the endeavour to build physics on facts, and the more deeply an idea depends upon the nature of things, the more rigid must the requirement be, not only that the facts permit the explanation but that they necessitate it. Chemical facts since the time of **Dalton** have all been in favour of the view that atoms do remain as such in molecules, although at first this was held only with more or less greater probability. First of all there was the fact that when a compound was decomposed, exactly the same atoms, both in kind and number, appeared again ; in addition to this there was the principle of the conservation of mass. This was found to apply to all chemical reactions, however much human skill endeavoured to improve the technique by which they were studied. Besides, there are the other so-called additive properties of matter. To-day we find ourselves more favourably situated, for we are able to recognize in an atom characteristics of such a kind that they will distinguish it from all others ; these characteristics repeat themselves both qualitatively and quantitatively in the compounds which the atoms form. We refer especially to X-ray spectra, and the phenomena of radioactivity. The X-ray spectra have already been discussed [*cf.* pp. 361 and 369].

Radioactivity is, as was shown on page 313, a process which only takes place in the atoms themselves. That this is so is obvious from the fact that radium emanation is a monatomic rare gas. At – 71° *C.* the emanation freezes and passes into the crystalline state. But even at the temperature of liquid air, which is more than 100° below this (– 186° *C.*), **Rutherford** was able to show that the radiation was still the same, although he continued to observe it for one hundred days, at the end of which period the quantity present had been reduced to $\frac{1}{200}$ millionth part of that initially used. **M. Curie** and **Kamerlingh-Onnes** found that the above remarks applied even when the emanation was investigated at the much lower temperature of liquid hydrogen (– 255° *C.*). Hence it follows that the atoms of this inert gas remain unchanged even when the emanation is present in the crystalline state.

Similarly pure metallic radium exhibited the same radioactive phenomena as the radium bromide from which it had been extracted. Further, it is known that all radioactive phenomena are not influenced by chemical reactions, so that after these phenomena had become established facts, it was proved that the atoms in molecules do persist as such, with all their characteristic properties.

2. But now that we have obtained such a variety of facts throwing light on the construction of an atom from a nucleus and electrons, we can examine this question a little more closely. We have seen that not all those phenomena which we have to ascribe to an element are essential to the element in the same way. The electrons can be separated from the nucleus,—the outer ones very easily, but the inner ones only by deeply penetrating cathode rays. But so long as the nucleus remains intact the whole atom can be restored. We therefore have to consider the nucleus as the really essential part of an atom, since it is responsible for the phenomena which arise in the mantle of electrons around it, in so far as it assists in the arrangement of these electrons in their own proper positions.

Even after a chemical reaction, the nucleus of an atom remains unaltered. As a proof of this statement we may cite the phenomena of radioactive substances. The conservation of the electron shells nearest to the nucleus is revealed to us by means of X-ray spectra; but that important changes take place with the valency electrons during chemical reaction is shown by the complete modification of the optical spectrum under these conditions.

The starting point for further considerations, for which our thanks are due chiefly to **Kossel**, **Lewis**, and **Langmuir**, is the connexion which exists between valency and electrolytic phenomena [*cf.* p. 273]. Those elements whose electron shells are completely closed, *i.e.* the inert gases, do not form ions in solution. The alkali metals, which occupy positions in the periodic system next to these gases, easily give up the one electron which is normally present in their outermost shell. The alkaline earth metals, which follow these alkali metals, have two such electrons which they lose in water, thereby producing ions with a double positive charge. On the other hand, those elements which precede the inert gases, viz. *F*, *Cl*, *Br* and *I*, form negative ions, by attaching to themselves another electron. The elements which occur two places in front of the inert gases, and which lack two electrons in their outermost shells, easily capture these, and form doubly charged negative ions. If we consider these facts more closely, it appears that all these ions have an electron arrangement similar to that of the corresponding inert gas, when they have either captured the electrons which were lacking, or parted with those of which there was an excess. Those elements of the middle group which are equally distinct from two different inert gases can, according to the external conditions, either become positive ions by giving up electrons, so that they simulate the next lighter inert gas, or they may become negative ions by capturing electrons, when they resemble the nearest but heavier inert gas. Accordingly, it appears that there is, in Nature, an endeavour for elements to form closed electron shells. In accordance with this remodelling of the electron arrangement, we have the fact that the ions have become similar to the inert gases in their chemical behaviour, for they have quite lost the power of participating in those reactions which were characteristic of the neutral atom. For example, the sodium ion is no longer able to decompose water.

When sodium chloride is formed from the two ions Na^+ and Cl^-, we have to assume that these two oppositely charged ions have been attracted towards each other, and that the attractive force then holds them rigidly together. In so doing each component has acquired the completed electron arrangement characteristic of an inert gas, so that the whole molecule only possesses closed electron shells. Such molecules which consist of several ions are called ionic molecules, in contrast to the atomic molecules. It is assumed that for these atomic molecules the electrons, in the shells which were originally completed, continue to revolve around their own nucleus, whereas the electrons which were initially in incomplete shells, form a new shell and revolve about the double nucleus.

We shall not refer to the attempts which have been made to describe the details of this picture more fully. Though we are still uncertain about the position and shape of the electron orbits in an atom, it is still more true that our knowledge of the arrangement of the electrons and nuclei in molecules is even more scanty. Since, however, molecules do possess spectra which are peculiar to them, the investigation of which during recent years has given us an insight into their structure, we have reason to believe that, when these investigations have been extended, much new information will be forthcoming.

In conclusion it must be admitted that the progress which has been made in science during the last few decades has been very remarkable, both for its extent and its depth. But these new ideas have brought other problems in their train, which patiently wait to be solved. Although the swarm of electrons enshrouds a large number of unsolved riddles, our knowledge concerning the nucleus itself is still in its infancy. We certainly know that it is the seat of an enormous amount of energy, for from time to time some of these nuclei explode and eject heavy particles at a very high speed, as we see in the case of alpha particles. In fact, it has been calculated that this velocity could be due to the repulsive force which exists between the alpha particle which has been released and the remaining portion of the original nucleus. If it originates at a distance of 10^{-12} cm. from the nucleus, which for this purpose we assume to be a point, the energy is then actually sufficient to impart to the alpha particle its high velocity. But if it is so, the question as to why the nucleus of a helium atom can exist at all becomes more difficult. Further, how the velocity of a beta particle originates is a problem which is more difficult to fathom, for although it is held back by the nucleus on account of its negative charge with almost the same force, it is nevertheless sent out with a velocity which approaches very closely that of light.

PART IV

THE PHYSICS OF THE ÆTHER

INTRODUCTION

HITHERTO the object we have had in view has been to obtain information concerning those bodies which we generally term the material world. But there is a whole series of phenomena which are not bound to this material world, but which can even take place in those regions where no matter at all is to be perceived or detected. Even if they do occur locally, amongst the elements of our world, yet we have every reason to believe that those bodies which we have discussed so far are not or rather not alone, and even then not for preference, the means whereby these phenomena are propagated. In the first place, it is true that light phenomena have their origin in the usual elements, but they are transmitted through interstellar spaces, where we are convinced that not even the merest trace of gas exists. Amongst this class of phenomena must be placed electric and other waves, now that their nature has become known. Strictly speaking, the force of gravitation also falls in this category. But it is unfortunate that our knowledge of its propagation through empty space is such an absolute blank, that it can very well be omitted from our present discussions.

The manner in which this question is discussed is quite unique. We do not know of any manifestation whatsoever of the æther from which we should be able to draw any conclusions concerning its properties. But since we have seen that light, electricity, and even gravitation, are transmitted across empty space, we conclude that this space is not really empty, but that it is filled with a material something, which supplies the means whereby these processes are propagated. The æther has therefore, for our purposes, no other properties than those which are necessary to explain the observed phenomena. The task which confronts us, then, is to show whether, from all these necessary properties, we can obtain a picture which shall at least meet all requirements without contradicting them.

CHAPTER I

THE UNDISTURBED PROPAGATION OF WAVES IN THE ÆTHER

1. The Rectilinear Propagation of Light

THE first fact to be ascertained concerning light, since only the most superficial observations were necessary, was that it is transmitted in straight lines. On a small scale, this is demonstrated by the sharp shadow cast by an object when illuminated by a point source ; on a larger scale, the orbits of the planets have been calculated on the assumption that light travels along straight lines, and on no occasion has any contradictory evidence been found. Confirmations of this hypothesis are both numerous and exact. Even if light were propagated in a direction which only differed very slightly from that of a straight line, the deviations over the enormous distances with which astronomy deals would be so great, that the fact that they have not been found must banish all doubt from our minds concerning this subject. It is indeed a different matter if we wish to understand this rectilinear propagation. What is it, then, which is finally propagated in straight lines ? In order to answer this we shall have to obtain an insight into the mechanism of the propagation of light. The **Huygens-Fresnel** principle has given us the necessary information, so that when we have become acquainted with it, we shall be able to consider this question again.

2. The Velocity of Light

For many centuries it was not decided whether light required time in order for it to be propagated, or whether it moved instantaneously through space. It would appear that in ancient times the general assumption was that it travelled instantaneously. The experiences of daily life, and the experiments which were then possible, justified this view. **Galileo** devised a simple test in order to settle this question. Two observers, *A* and *B*, each furnished with a lantern, were placed at a distance apart of one mile. At a definite time *A* removed the cover which had previously hid his lantern from view. As soon as *B* saw it he was able to reply in the same way. The time which elapsed between the removal of the cover from the first lantern and the perception of the light from the other by the observer *A* was, according to Galileo, the time necessary for the light to travel from *A* to *B* and back again. The experiment led to no result for no interval of time could be ascertained. We now know that the experiment must necessarily be inconclusive, for the distance between the two observers was far too small : the time which light needs to travel across such a short distance could not be detected by the means then available. From this unsuccessful experiment it was naturally not possible for any definite conclusion to be made concerning the velocity of transmission of light. It could only show that the velocity is greater than that which was detectable by these means.

Descartes also believed that light was propagated instantaneously. The proof which he gave for this view is very remarkable. If time is required for light from the stars to reach us, then the velocity of light, in conjunction with that of our earth, will impart to the fixed stars an apparent motion; the problem is similar to that of rain which, although falling vertically, appears to meet a train in which we happen to be travelling. This view is worthy of our attention, for two reasons. In the first place, Descartes' view was the reason that many contemporaries would not acknowledge the evidence of **Olaf Römer**, brought forward in order to show that the velocity of light was finite; and in the second place, the argument itself even contained the counter-evidence which was used by **Bradley** fifty years later (1728) in order to determine the velocity of light across interstellar space.

Römer's Determination of the Velocity of Light. The first determination of the finite velocity of light was really quite accidental. In 1675 the Danish astronomer Olaf Römer was engaged in observing the motions of Jupiter's satellites from an observatory in Paris. He discovered certain variations in the period of one of these satellites which could not be accounted for except by the fact that light must travel with a finite velocity. From observations on the period of rotation of one of these moons when the earth and Jupiter were on the same side of the sun, the position of this moon at any other future time could be predicted. Römer calculated the times at which one of the moons should just pass behind the planet as seen from the earth, and then actually observed these disappearances on future occasions. The discrepancies between the calculated and observed times became greater as the earth moved away from Jupiter.

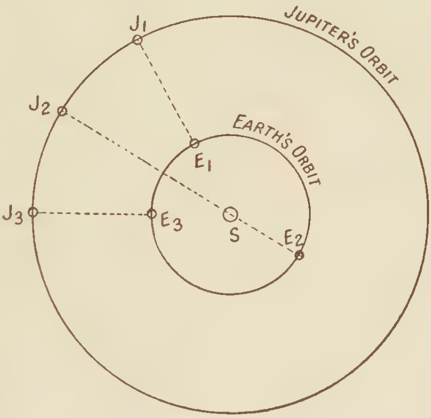


FIG. 166.

This difference attains a maximum value when the earth and Jupiter are on opposite sides of the sun. Let us suppose that when E_1 and J_1, Fig. 166, represent the positions of the earth and Jupiter respectively, the period of revolution of this one moon is first determined. At some future time (about seven months) the earth and Jupiter will be at E_2 and J_2, *i.e.* they will be in opposition. In travelling between J_2 and E_2 the light will traverse a greater distance than in coming from J_1 to E_1: this difference will be equal to twice the earth's orbital radius. If light has a finite velocity, the observed disappearances will thus occur later than was calculated, the difference between the observed and predicted times being equal to that taken by light to travel across this increased distance.

The diameter of the earth's orbit is 300×10^6 km., whilst the delay amounted to 1000 secs. The velocity of light is therefore 300,000 km. or 186,000 miles per second.

At a later date (1728) **Bradley** measured the velocity of light from the so-called aberration of light as manifested by the fixed stars—it will be remembered that Descartes had originally introduced this idea. In

1849 **Fizeau** determined this constant from observations made across a distance of 8·6 km. in air.

Foucault's Method for Determining the Velocity of Light. The experiment which **Foucault** carried out, also in 1849, but this time in a room of ordinary dimensions, was destined to play an important rôle in the study of light. Light, passing through a narrow slit *L* (Fig. 167), fell on a mirror *M* which could rotate rapidly about a perpendicular axis. At a distance of 4 inches a fixed mirror was placed at *A*. When the mirror *M* stood still the light fell on *A*, from whence it was reflected to *M* and then back towards *L*. In front of *L* was situated a transparent glass plate *G* inclined at an angle of 45° to the direction *LM*. A portion of the light returning to *L* was therefore brought to *B*, where it could be observed. By means of lenses suitably designed and placed, a well-defined image of the slit was obtained at *B*; they have been omitted from the diagram since they are not essential to the present argument. When the mirror was in rapid rotation the image no longer appeared at *B*, for during the time in which the light went from *M* to *A* and back again, the mirror had made a small rotation. In consequence of this the image now appeared at *B′*. By measuring this small shift *BB′* the angle through which the mirror *M* had rotated could be calculated; combining this with the number of revolutions made by *M* per second, the time for the light to move across twice the distance *AM* was obtained. The velocity of light could then be deduced at once.

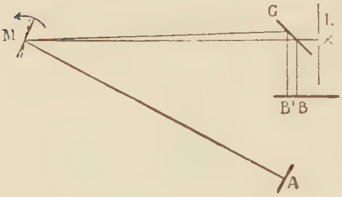


Fig. 167.

In some later experiments by Foucault the above distance *MA* was occupied by a tube, closed at each end by means of a piece of plain glass, and completely filled with water. The shift was now greater since the light required a longer time to pass through water. The shift was increased in the ratio 4 : 3. It therefore follows that light travels in water with a lower velocity, in fact, it is only 225,000 km. per sec. in water.

The Velocity of Electromagnetic Waves is measured by some other method in which use is made of the relationship that the velocity of propagation, c, is equal to the product of the wave-length λ, and the frequency ν, *i.e.* $c = \nu\lambda$ [*cf.* equation (43)]. In the first place, waves of a definite frequency ν were therefore obtained. This number depends upon the dimensions of the apparatus—its capacity and self-inductance—according to the formula $T = 2\pi\sqrt{LC}$ [*cf.* equation (159)]. Since C and L can be ascertained from electrical measurements, in these experiments T, or what is equivalent, $\nu = 1/T$, can be considered as a known quantity.

It is still necessary, however, for λ to be determined. This was done by producing stationary waves, as in the case of waves in an elastic medium, by allowing interference to occur between the advancing waves and those which had been reflected. The distance between two of the series of nodes so excited is half a wave-length. After he had corrected a small mistake, the final value found by **Hertz** for the velocity of electromagnetic waves was 300,000 km. per sec., or 3×10^{10} cm. per sec.

The Velocity of X-rays. We must now refer to the experiments by which the velocity of X-rays has been determined. **Marx**, to whom these experiments are due, compared the velocity of X-radiation with that of electric waves in wires and found them to be equal. It is not possible, however, for us to consider the details of these experiments.

3. THE NATURE OF THE TRANSMISSION OF LIGHT

The question concerning the manner in which light is propagated is, in general, very intimately connected with the problem of the nature of light itself. For the transmission will be essentially different according as to whether light consists of particles projected at a high speed, or of a progressive wave motion. Even in classical times the phenomena of light attracted the attention of the Greek philosophers, but no definite ideas concerning the nature of light were expressed. They imagined that the process of seeing was due to bodies being touched by rays which were a kind of feeler sent out from the eye. The only ancient philosopher who stands out pre-eminently above all others in dealing with this problem is **Aristotle**. He recognized quite clearly that vision was not due to something sent out by the eye, but must be assumed to have its origin in the body which is seen, *i.e.* light is transmitted from the body to the eye in a similar way to which sound comes from its source through the air to our ears. He required a suitable medium for the transmission of light, and certainly believed that its velocity through the medium was infinite. On one occasion he wrote to Plato, ‘ If the eye sends out feelers, as a lantern emits light, then why is it that we cannot see at night ? ’ But in the centuries which followed little heed was paid to these ideas. The Arabian **Alhazen**, who lived about 1000 years later in Spain, was the first to apply himself to the study of Aristotle’s doctrines again. He was even able to supplement them in many ways ; in particular, he maintained that the velocity of light was finite. But even the progress which he made was only just known by his contemporaries, and even Descartes ‘ proved ’, as we have already seen, that light travelled instantaneously. The first detailed theory of light originates from **Newton.**

Newton’s Emission Theory of Light (1669). According to Newton, light consisted of very small particles sent out from the luminous body. When these particles enter our eye they produce the sensation we call sight. Now, according to the more advanced views of that time, this theory should be able to explain all light phenomena then known. To a great extent it was successful in this. The fact, for example, that a pane of glass allows a portion of the light incident from the sun to pass through, whilst another portion is reflected, was explained by assuming that these particles were subject to ‘ fits ’, in consequence of which some were transmitted whilst others were reflected. It might perhaps be imagined that these light corpuscles were not round, but angular, and that they rotate as they move along. Those particles which strike the glass with one of their angular points will pass through, whilst those whose faces collide with it will be reflected.

Huygens’ Wave Theory of Light. Even at the same time, another theory of light was being proposed by **Christian Huygens**, a Dutchman. It is expounded in the much admired small work, ‘ *Traité de la lumière* ’ (1690). In it light is represented as a motion in an all-pervading medium. The chief reasons which led him to propound this theory of light were as follows.

1. The idea with which he was already familiar, that heated bodies and those rendered luminous by heating, contain particles undergoing rapid motions.

2. The fact that several light rays can pass through one another and intersect, without any disturbance to the paths taking place.

3. The fact that the velocity of light had been found to be independent of the intensity of the light radiations, whereas if the emission theory is adopted, the intensity of light can only be conceived as due to an increase in the velocity of the light corpuscles.

Huygens had even published his theory in 1678, but his contemporaries would accept no other view than that of Newton ; they considered Huygens' work to be insufficient, and so Newton's emission theory prevailed for more than a century. It is true, however, that several phenomena presented great difficulties, when an attempt was made to explain them in terms of this hypothesis, and with the lapse of time these difficulties increased more and more. But even Huygens' theory had its difficulties too, and they were only solved by partly changing some of his original assumptions. In particular, Huygens had carried the analogy with sound waves too far when he assumed that light consisted of longitudinal waves. **Fresnel** first assumed that they were transverse, and thereby obtained a natural explanation of the facts then known. Many years elapsed, however, before Newton's views were abandoned. This happened finally in the year 1849, in consequence of **Foucault's** determination of the velocity of light in different media, to which reference has already been made. The reasons for this were as follows.

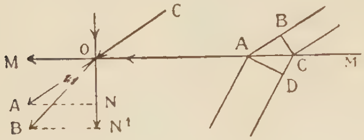


FIG. 168.

If *CO* is a ray of light, Fig. 168, incident upon *MM*, the surface of a new medium (let us say water), then when it has passed into this medium it becomes deflected towards the normal *ON* ; instead of passing from *O* to *A*, it travels, on the contrary, towards *B*. The two theories of light, the emission and the undulatory theories, had to explain this fact. Newton proceeded thus. The velocity of the incident particle can be resolved into two components, one perpendicular to the plane of separation of the two media, the other tangential to it. The second component is merely directed along the surface of separation and therefore is not changed on entering the second medium. The other component velocity must therefore be so changed, that the resultant path is consistent with that actually observed. This can only happen if this component velocity is greater in the new than in the old medium. Newton therefore assumed that when a particle approached such another medium it is attracted by it and so moves more quickly, the perpendicular component of its velocity becoming greater (*i.e.* $ON' > ON$) ; a deviation is then produced, as is actually observed in practice. This explanation stands or falls, according to whether the velocity of light in water is greater or less than that in air.

The representation of this same process in terms of the general theory of waves has already been detailed [*cf.* p. 86]. For the sake of comparison the diagram has been reproduced here. According to the wave theory, whilst light has been moving along a distance *BA* in the first medium, the distance *CD* has been accomplished by light in the second medium. The refraction takes place because the distance *CD* is shorter than *BA*. A decision as to the validity of these theories of light therefore depends upon the reply which experiment gives to the following query : Does light travel in water more rapidly or more slowly than in air ? To answer this question was the primary reason why

Foucault devised his experiment, the result of which, as we have already learnt, was in favour of the wave theory. From now onwards all phenomena will have to be explained in terms of the wave theory.

The rectilinear propagation manifests itself in accordance with the fundamental law of Huygens and Fresnel, which we discussed when dealing with the theory of waves in general. For the time being it will suffice to refer to the explanation given on pages 81 *et seq.*

4. THE MEDIUM IN WHICH LIGHT TRAVELS

The Elastic Æther. The attempt to fathom the process of the transmission of light leads at once to a discussion of the medium which is to carry this wave motion. Now it is absurd to think of a wave without a material something in which the wave motion is executed. It is therefore necessary to assume that the whole of interstellar space, like that between the individual air molecules, is filled with a material something which picks up the waves and carries them forward. This medium must be endowed with such properties that it is able to fulfil this task ; at this point difficulties begin to confront the wave theory. The phenomena of polarization prove, in the first place, that light consists of transverse waves. Now transverse waves can only be propagated through a medium in which the various movable particles are constrained together, so that when one is moved the neighbouring particles are pulled to one side, as we have already described [*cf.* p. 74]. But where such forces exist, they must give to bodies a fixed shape. We therefore have to consider the space between the stars as well as that in our own atmosphere, in which we move and breathe, as completely filled with a rigid substance. But we do not notice it as we move through it, and the heavenly bodies, which, on the whole, are moving with enormous velocities through space, also appear to experience no resistance as they move through this all-pervading substance. If we are to admit of a new element which is to occupy all space, then it could at most only be a rarefied gas, which would not offer the least resistance to a motion through it. But the propagation of light waves, which is the main task to be fulfilled by the æther, imperatively requires an elastic and therefore a solid medium.

Science was here confronted with a very great difficulty. It was not as if the theories contradicted one another, for that would not be any great difficulty. It had long been learnt how to make a sharp distinction between theories, which are the work of men, and which, on the morrow, may be replaced by another, and facts, which irresistibly compel our recognition. Man wishes to do justice to facts ; but he is himself bound to certain laws of thought which demand attention and recognition, just as definitely as do the other facts. But here it appeared as if these laws of thought were in opposition to the facts, and therein lay the magnitude of the difficulty. The existence of transverse waves appeared to be a fact, and that the heavenly bodies, like our own, can move without any appreciable opposition through space, likewise appeared to be a fact. Our laws of thought require a medium having rigidity, whilst another fundamental law of thought tells us that it cannot be both rigid and gaseous at the same time, as the facts would seem to indicate. This difficulty has remained ever since without being solved ; it appeared to be quite insoluble. Such a condition is something we cannot tolerate. We can only wait patiently, hoping that soon a way out will be found ; but the longer we wait without a solution being discovered, the more obvious does it become that something is wrong with the fundamental assumptions upon which we have based the whole argument. Similar instances have occurred on many later occasions; they are always of the utmost importance, for the more difficult it is to find a solution, the greater must be the progress made when the riddle is eventually solved.

The Electromagnetic Æther. The problem of the nature of the æther has not even yet been solved; but, fifty years after Fresnel's time, it was first directed into somewhat different channels. Maxwell's electromagnetic theory proved that light waves are not elastic, but electromagnetic, so that it was no longer necessary for the æther to be able to transmit elastic waves, but it must be capable of propagating waves which are electromagnetic. This implies that it is capable of becoming polarized (electrically) [*cf.* p. 243]. Electric particles must be found in it which are capable of undergoing a small displacement; but forces must then come into play which will cause them to return to their zero position. One can now no longer say that such properties require a rigid body. In this respect the electromagnetic theory helped to relieve the situation somewhat, so that for this reason it was welcome with great joy. According to this theory the small electric charges are in rapid vibration, but since they are constrained to a zero position, forces arise when they are displaced from this position which tend to make them return; the magnitude of these forces becomes greater as the displacement is increased. At that instant when the motion is reversed the velocity is zero, the particle therefore having no kinetic energy,—the only energy which it possesses is electrical, due to the equilibrium having been disturbed. But so long as the carrier of the electric charge is in motion, this moving charge is equivalent to an electric current so that a magnetic field is associated with it. The electromagnetic energy which is therefore present will attain its maximum value when the current is greatest, *i.e.* when the velocity of the charge is highest. This happens at that moment when the charge is passing through its position of equilibrium, *i.e.* when the electrical energy is zero. The case is exactly similar to that of an oscillating pendulum, the energy of which is continually changing between energy due to position (potential), which is all the pendulum possesses at the extremity of its swing, and kinetic energy, which is the only form of energy present when the pendulum moves through its zero position. In the present problem the potential energy is replaced by electrostatic energy, and the kinetic by magnetic energy, so that the self-inductance will play the same part here as inertial masses do in mechanical vibrations.

Some Remarks concerning Changes in Theoretical Physics. Perhaps many readers, who are not satisfied with this change in our theory, will ask how it is possible for us to discard the idea of elastic waves and replace them by waves which are electromagnetic, when the first theory appeared to be supported by facts, and it has always been maintained that facts possess a certain degree of finality. Now is the time when an answer to this question will be most appropriate. Facts are certainly landmarks which science cannot efface, and they have to be heeded without reserve both in the old and new theories. For example, in the present instance, we had studied the phenomena of interference and polarization, had measured wave-lengths and the velocity of propagation, and found that all these were embraced by one theory, viz. that of elastic waves. This hypothesis lay very near to the truth; in fact, at that time, it was the only possible theory, since no other waves were known, except those due to the elasticity of the medium in which they travelled, so that no others were considered possible. But the progress which was subsequently made in reference to electricity showed that there were other types of wave which were also capable of transmission through space. This indicated that the conclusion which had been made from observed facts, that light waves were elastic in nature, was no longer the only possible one and therefore lacked finality. This has always happened whenever earlier views have had to be discarded in the light of further research.

Let us remind ourselves of the decision given by scientists against **Prout's** hypothesis, which had stipulated that all elements were built up from hydrogen. Their rejection appeared to be quite valid, for several elements had atomic weights which were not integral multiples of that of hydrogen, so that a hydrogen atom could not be the smallest particle of matter from which all other elements were built. The idea that the atomic weights were possibly only mean values was so remote that it was never expressed, for all known elements behaved in all their actions as perfect entities, so that it had been concluded that atoms of one element were all alike. The opposite view appeared to be inconsistent with the facts then known. It was only much later that such critical thinkers as **Crookes** (1879) called attention to the precariousness of this conclusion. But no attention was paid to these objections since they appeared to be so improbable, and from what was then known about the chemical behaviour of the elements, they were improbable. It was only when further progress had been made in natural science that it came to be realized that it is not the atomic weight which determines the properties of an element, and that atomic weights expressed by fractional numbers are really average values of integral numbers. Prout's hypothesis was no longer contradicted by the fractional values of some of the atomic weights. We see, therefore, that it is always the limitations of the human mind which compel us to make supplementary changes in our theories. After these experiences there is no need at all for us to frame theories which will apply less rigorously to the facts, for it has been shown time and again how cautiously one must proceed when the aim is to draw conclusions from the facts, and then use them to construct a theory. Such experiences should rather warn us to be cautious, but they need not cause us to despair of the activity of our mind, for even the fact that we have recognized the inadequacy of former conclusions and replaced them by better ones, only serves to prove how the human spirit strives for truth, and how steadfastly, even if only at the cost of great labour, the truth is more nearly approached.

As regards the actual carriers of light waves, we shall return to discuss them in the concluding pages of this book, when we have become acquainted with all the facts which bear upon this subject.

CHAPTER II

THE DISTURBED PROPAGATION OF LIGHT

WHEN a light wave is incident upon a solid or liquid body, the direction in which it travels is, in general, changed considerably. A portion of the incident light is reflected from the surface of the body, another portion penetrates into the inside of the body where it is absorbed, whilst the rest eventually appears at the other side of the body if this is transparent. The magnitude of one or other of these components varies from one body to another, and, moreover, it depends upon the wave-length of the incident light. We say that glass is transparent because it transmits most of that light falling upon it which is involved in vision. But other waves, *e.g.* ultra-violet light, are absorbed, and if our eyes were sensitive to such radiations, then glass would not be transparent, so that it could not be used as a window.

From these facts it follows at once that even if material bodies are not the real agents which carry a light vibration forward, they certainly do participate in its propagation, when a light ray passes through the body. We shall be able to observe the way in which this takes place on many occasions in this present chapter. For the time being it suffices for us to know that bodies do influence the propagation of light and therefore must participate in the process whereby it is transmitted.

1. The Reflexion of Light

Let us consider first of all that portion of light energy which is reflected at the surfaces of bodies. We assume that we have a flat surface of water, or better, of mercury. It is well known that such a surface can be used as a mirror. What does that mean ? It means that if we look at that surface we shall see the objects round and about us, as if they were in the liquid. We see our own faces and everything around us. In other words, light comes from the surface of the liquid into our eyes, and this light influences them in exactly the same manner as **the light sent out by these same objects when we observe them directly without the aid of a mirror.** This can be explained very easily in terms of the theory of wave motion. It has already been given in connexion with the general theory of waves [*cf.* p. 85], and the law, which was there deduced quite generally, is valid for the present problem concerning a plane reflecting mirror, when light falls upon it. The light wave is reflected back from those points on the surface at which it is incident, as if it originated from a point lying exactly as far behind the reflecting surface as the actual source lies in front. The distance of the light source from the eye is also indicated, but not directly. That which excites the sensation in our eye—called sight—is only what, with the wave and in the wave, acts upon the eye. But the distance of the source is really included in the curvature of the wave front, for this becomes smaller as the distance increases. Each small plane element of the reflecting surface gives rise to a portion of the reflected wave, by means of which the curvature and consequently the position of the source is completely defined. For example, that portion of the wave which our eye receives (Fig. 44, p. 85) originates from the small area BB_1, and it is by means of this portion of the wave surface that our eye observes the radiant point L_1. For this portion of the wave, it is true to say that the angle of incidence is equal to the angle of reflexion.

Reflexion at Rough Surfaces. If a surface is not plane, then we imagine that all the tiny elements of area, which go to produce that surface, are all orientated in different directions. Then the light which falls upon the various elements will be reflected in various directions. An eye which is directed upon the surface will perhaps receive from one such element a large portion of the light which it reflects, for another neighbouring one nothing at all, whilst from a third only a little may be received. Such surfaces are said to be rough. Let us assume that a point source of light illuminates such a rough surface. In the former problem concerning a plane surface, an eye, which we assume to be quite near to the source and which is directed towards the surface, receives a certain amount of light energy ; it comes from the small area f (Fig. 169). If F is the area of the pupil of the eye, then that of the surface from which the energy received originates is given by $f = F/4$, since the diameter of f is only one-half that of F. If, now, the surface is rough (Fig. 170), then the light from f is no longer all reflected in the same direction ; on the contrary, for a perfectly rough surface it is distributed in all directions upon an hemisphere described in front of the surface. If d is the distance of the eye from the point A, then the reflected light is distributed over an area $2\pi d^2$, so that only the fraction $F/2\pi d^2$ is received by the eye. It is true that the eye now receives light from other portions of the illuminated surface, but this will be incident upon different parts of the retina. The resultant effect is that we see a greater area of the surface, but the impression from A will not be altered thereby. We therefore no longer receive a light impression which is similar to that which we should receive from a smooth mirror or surface. In the case of reflexion from a smooth surface a light wave which falls upon the eye has its centre or origin in L_1. The centre of a wave starting from a rough surface is A, *i.e.* it is in the surface itself. That is why, in the first case, we see L_1 and in the second A.

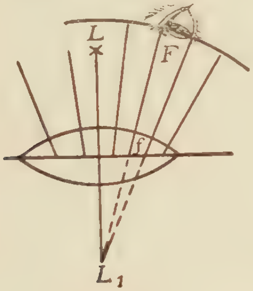


FIG. 169.

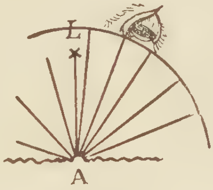


FIG. 170.

Reflexion at a Curved Surface. We now turn to consider the reflexion from surfaces which are not rough nor yet plain, but, on the contrary, those surfaces in which neighbouring elements of area have approximately the same orientation in space. This not only applies to cylindrical surfaces, but, in general, to all surfaces which can be represented by a simple mathematical expression. Here we shall only consider the simplest case of all, and that is the reflexion of light from a surface which is spherical.

On account of the repeated application which we shall make of a certain property of a circle, it will be convenient to remind ourselves of it at this stage. If $2h$ is the length of a chord in a circle whose radius is r, Fig. 171, and e the perpendicular distance of the chord from the circumference, *i.e.* it is the height of the minor segment of the circle, then $h^2 = e(2r - e)$. In the particular problem when h is small, e can be neglected in comparison with $2r$, so that the equation then assumes the more simple form $h^2 = 2re$, or, as we shall often use it,

$$e = \frac{h^2}{2r} \quad . \quad . \quad . \quad (188)$$

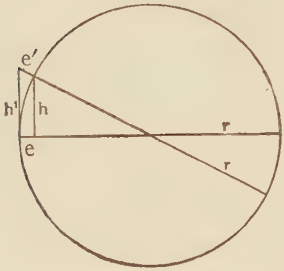


Fig. 171.

If a tangent h' to the circle is then drawn and its end point or terminus joined to the centre, then $h'^2 = e'(2r + e')$, where e' is the perpendicular difference between the terminus of h' and the circumference of the circle. If h' is assumed to be small, e' can again be neglected in comparison with $2r$. Hence, we also have $e' = h'^2/2r$. Comparing the values of h and h', we have $h : h' = (r - e) : r$. Now if h is put equal to h', we have with the same degree of accuracy $e = e'$. These equations are certainly only true for values of h and h' which are essentially smaller than those indicated in Fig. 171.

We now return to the consideration of a light wave reflected from a spherical surface. Let us assume that O, Fig. 172, is a **luminous point** or **radiant**, *i.e.* a light source of very small dimensions. At first, let O be further removed from the spherical surface S_1S_1 than is its centre of curvature C. Then a wave sent out from O will first meet the mirror along its periphery. When this wave meets the pole P of the mirror, then in the absence of the mirror the wave front would be at APA_1. Actually, however, it is not there, and the further treatment of the problem is similar to that used when discussing reflexion at a plane surface. The portion of the wave to the right of the mirror (shaded in the figure) is actually not there but lies to the left of the mirror on account of the reflexion which has taken place. This is constructed by making use of the law which states that the angle of incidence is equal to the angle of reflexion. The wave front DPD_1, which is thereby obtained, has a curvature which is more pronounced than that of the original wave. This is the shape of the wave formed after a complete reflexion and which gives rise to the image I. It is obvious that the curvature of this return wave front is greater than that of the spherical surface, and that its centre is nearer than C to S.

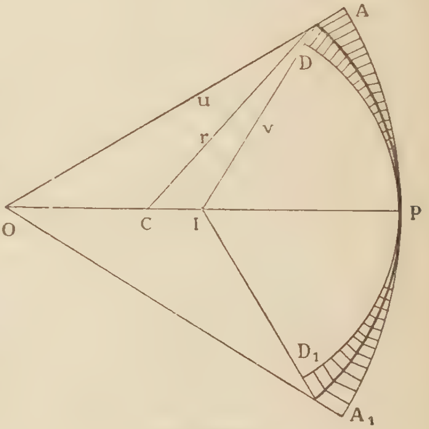


Fig. 172.

In the second case we assume that the radiant O is nearer the mirror than is C, its centre (Fig. 173). The incident wave front will now first meet the mirror at its pole P, and by the time that the wave has reached the periphery of the mirror, the central point of the wave would have arrived at E if the mirror were absent. Under these circumstances the crescent AEA_1PA would have been formed; actually, however, this portion of the advancing wave forms another crescent on the left-hand side of the mirror. The radius of curvature of this wave front DI is greater than that of the mirror. The points I in Figs. 172 and 173 are called the images of the objects O: they are produced by the action of the mirrors.

Let us now consider the formation of this image somewhat more precisely. We notice that in the two examples described above the whole energy of the wave is brought together at the image I. Such an image is said to be **real** in contrast to one which is **virtual**. This latter type of image has already been met in the image formed by reflexion at a plane surface. On that occasion the rays did not actually proceed from the image L_1 to our eye [*cf.* Fig. 44], but, after reflexion, their direction was such that they appeared to come from L_1. When a piece of white paper is held at the position I on each occasion, very different results are obtained. When the image is real we see that light energy is radiated from the position where the image is formed, and that the image is exactly similar in colour and shape to that of the actual object; in the case of a virtual image nothing at all is observed when the paper is held at I.

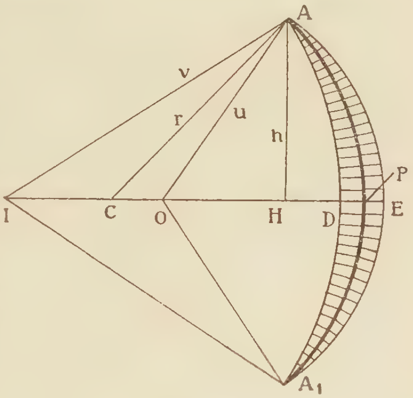


FIG. 173.

First of all we shall consider the formation of a real image. From the actual object O, which we assume to be a luminous point, spherical waves originate which are transmitted through the region round about. When a mirror (or later, a lens) is interposed, the light energy is distributed over a new wave front which does not travel outwards in a direction from the centre to the circumference but proceeds from the circumference towards the centre. Then, when all its energy has become concentrated at its new centre, another luminous point is formed; this is the image of the first point and its energy is the same as that which was sent out initially.

The Effect of Phase. Even this, however, is not all. When a certain quantity of light energy moves from a spherical surface towards its centre, it does not follow that on all occasions a luminous point will be formed. Let us contemplate the case in which all the waves do arrive at the centre but that their phases are then different. If one portion of a wave makes a point move upwards at the same time as another urges it downwards, then the two actions will oppose one another or even destroy one another completely. It therefore follows **that an appreciable amount of light energy can only become concentrated at a point when all portions of the arriving wave are in phase with one another.** Now the phase is identical at any points in the original wave front, since each portion of the disturbance was sent out from the source O at the same time. Moreover, since the same number of vibrations are executed in the same time, and the disturbance during each vibration has moved a distance λ, the phase will be the same at all points equally remote from the source.

If O is the source of light, then all points on any arbitrary sphere whose centre is O, will be in the same phase. Thus points on the wave APA_1, Fig. 172, will all be in phase. Now the wave surface after reflexion has been constructed by making distances to the left of the mirror equal to the corresponding ones on the right of it, so that there is equality of phase at all points on the wave front DPD_1 [or ADA_1, Fig. 173]. If we could therefore prove that all points on the reflected waves in Figs. 172 and 173 lie on a sphere, then the disturbances would all reach the centre at the same time, and having moved across equal distances, they would all be in phase. Under these conditions they would all act together and form an image, which would then be the seat of an intense wave motion.

Let us use Fig. 172 for the purposes of our proof. We therefore have to show that all points on ADA_1 lie on a spherical surface, *i.e.* that all points on this surface are equally remote from the image T. In the first place we choose the three points A, A_1, and D, on this wave front, and through them draw a circle, the radius of which is easily calculated. Let this radius be v; the reader must appreciate that we do not assume this circle to coincide with the wave front. Let a be the radius of the wave front whose centre is O. Now PE was made equal to PD in constructing the reflexed wave. With the aid of the lemma proved above, we have

$$DP = HP - HD = \frac{h^2}{2r} - \frac{h^2}{2v},$$

and

$$EP = HE - HP = \frac{h^2}{2u} - \frac{h^2}{2r}.$$

Since these two expressions are equal, we have, when $\frac{h^2}{2}$ is eliminated,

$$\frac{1}{v} + \frac{1}{u} = \frac{2}{r} . \qquad (189)$$

From this equation we obtain a value for the radius v. That is, the circle whose centre is equally distant from the points ADA_1. If now, instead of A and A_1 any two other points on the wave ADA_1 are taken, then the radius of the circle passing through them and D can again be calculated. The question is whether this second circle is the same as the first. If it is, then all points on the wave front lie on a spherical surface whose centre T is equally distant from all points on it, and the waves arrive there in phase.

For these other points the calculation will be exactly as before, except for the different numerical value of h. But we have seen that this quantity disappears from the final equation, so that the value for v does not change even when h is altered. We have therefore proved that we obtain for all points on the wave surface the same centre, *i.e.* they all lie on a spherical surface. What we have proved for this cross-section of the wave, holds for all other cross-sections. Hence, the centre I is not only that of the arc shown in the diagram, but it is the centre of all arcs which are the intersections of other planes and the spherical surface.

The proofs for the other cases which arise must be left as an exercise to the reader, now that the above has been worked out fully. The formula (189), or if we put $f = \frac{r}{2}$, is the well-known formula applicable to spherical mirrors, viz.

$$\frac{1}{v} + \frac{1}{u} = \frac{2}{r} = \frac{1}{f} \qquad . \quad . \quad . \quad . \quad (190)$$

If u is very great, which is the case when the light comes from the sun, then $\frac{1}{u}$ can no longer be distinguished from zero. Hence, for light coming from a source at an infinite distance, *i.e.* so-called parallel light, we have $v = f = \frac{r}{2}$. This means that parallel rays are brought to a point at distance $f = \frac{r}{2}$ from the pole of the mirror. This point is called the **focus** of the mirror, and the light from the sun arriving at that position is able to raise small objects to incandescence.

A glance at the above formula shows that the equation is not altered when u and v are interchanged, *i.e.* if I is the image of an object at O, then O will be the image of an object placed at I. Hence, when a luminous object is placed at I, its image will be seen at O providing the mirror is not changed. The two points O and I are said to be **conjugate foci.**

Spherical Aberration. In deriving the above formula we have availed ourselves of a development which is only approximately true. It is only when we limit ourselves to a small portion of the mirror that the formula agrees tolerably well with experiment. One can easily be convinced that this is so. If the mirror is covered up except for a small area around its pole a very sharp image will be obtained, since all portions of a wave give rise to an image at the same point I. When, however, a large area of the mirror is used to produce the image, this image will become somewhat diffuse, for all portions of a wave do not produce an image at exactly the same point. This defect of the image is known as spherical aberration ; it is produced because all the zones of the mirror do not give rise to an image at the same position as the central zone does.

Objects of Finite Dimensions. When we no longer have to deal with one luminous point but others which are adjacent to it, the above construction can be applied to each such point. Each luminous point is joined by a straight line to the centre of curvature, this line being produced until it meets the mirror. Since the mirror which we are contemplating is a portion of a sphere, all lines passing through its centre are equivalent, so that the same equation applies to points on them as for points on the axis of the mirror, only the mirror is not intersected by these lines at its pole. These lines are therefore called the **subsidiary axes** of the mirror in contrast to the first which is known as the **principal axis.** In fixing the position of the image this is all the less important, since our proof only involves small areas around the pole of the mirror, and such small areas can be considered equally well for both types of axis. The position of the image of a luminous object placed near to the principal axis is therefore given by equation (190). It will be found, however, that neighbouring points in the object give rise to point images which are also adjacent, but the distance between any two of them is, in general, not equal to the distance between the two corresponding points in the object. The ratio of these distances is called the **magnification** of the mirror. It can easily be shown that this ratio is equal to the ratio of the distances of the image and object from the pole of the mirror.

When it is not a question of understanding the formation of an image, but, as is frequently the case in practical optics, only of determining the position of the image, it is sufficient to determine the centre of the reflected waves in the shortest possible way. It is usual to determine that point where waves, reflected from a few parts of the mirror, are brought together. Strictly speaking, any portions of the wave could be used, but, for our purpose, it is better to consider a few definite portions on account of the simplicity which is thereby introduced into the work. For example, it is known that the whole of the wave coming from an infinitely distant source is brought to a focus at one point. Hence, any 'ray' which is parallel to the axis of the mirror will be reflected so that it passes through the focus. Conversely, a ray which passes through the focus before meeting the mirror will leave the mirror along a direction parallel to its axis. Further, every ray passing through the centre of curvature will meet the mirror at right angles and therefore be reflected back along its own path.

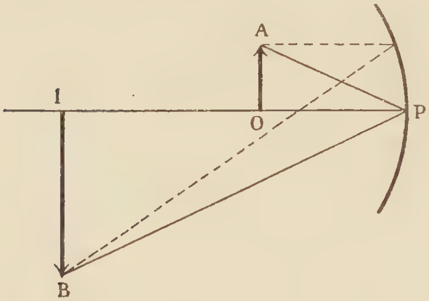


Fig. 174.

The magnification produced by a mirror is easily calculated by making use of the above statements. An object of height OA, placed at O, Fig. 174, has its image of height IB at I. In order to find the law of magnification, a ray is drawn from A to the pole of the mirror; it will be so reflected that the angle of incidence is equal to the angle of reflexion. The two right-angled triangles AOP and BIP are therefore similar: hence $OA : IB = OP : IP$. The ratio of the heights of image and object is therefore equal to the ratio of the distances of image and object from the pole of the mirror.

2. The Refraction of Light

When light passes from one medium to another, in general, the light is refracted. The relationships which exist have already been discussed [*cf.* pp. 86 and 388]. The refraction takes place because the velocities of propagation in the two media are different.

Let us endeavour to make ourselves clear on this point. The disturbance in the one medium, as in the other, passes from point to point so that the frequency must be the same in each case. For otherwise one of two adjacent points would be moving in a contrary transverse direction to the other, a condition which is quite impossible, since the vibrations at all points must be excited and handed on from neighbouring points, and must therefore move to and fro in step with them. But if the frequency of the waves is the same in each medium, then their velocity in one is given by $c = \nu\lambda$, whilst in the second it is $c' = \nu\lambda'$, *i.e.* $c : c' = \lambda : \lambda'$. The wave-lengths in the two media are therefore different, their ratio being equal to that of their velocities. It therefore follows that the same wave will travel across different distances in two media in the same interval of time; in fact, these distances are proportional to the velocities, or, according to the relationships already established [*cf.* p. 86], inversely proportional to the ratio of the refractive indices. If, therefore, two portions of a wave are in phase initially, they are also in phase when they have traversed paths in different media which are inversely proportional to their refractive indices. This is a very important principle, for the essential condition for the formation of an image is that all the waves which contribute towards it should be in phase. In lenses the rays of light are continually passing through different media.

Refraction by a Prism. A problem which is especially important in connexion with the study of optics concerns the passage of light through a prism. Essentially a prism consists of a transparent substance, bounded on two sides by inclined plane surfaces. The refractive index for the media between these two faces is different from that of the outside medium. Let us assume that a parallel beam of light travelling from left to right (one can imagine that such a beam has had its origin in the sun) falls upon a glass prism *ABC*, Fig. 175. *A* is the first point in the prism which the waves meet. The direction of the light in the prism can be found in the manner previously indicated. Now whilst the upper portion of the beam travels from *E* to *C*, the lower portion passes through glass with a smaller velocity, and in the same time arrives at *G*. This ray will continue to move in the same direction until it reaches *B*. Whilst it has been moving from *G* to *B*, however, the upper ray will have travelled a greater distance in air and have arrived at *F*. The advancing wave fronts will therefore be in phase first along *AE*, then along *CG*, and finally along *BF*. But even when the wave front is not entirely in one medium it will still be true to say that the phase is the same at all points on it at any one time. The emergent wave travels through air in a direction perpendicular to the wave front *BF*. The figure shows that each time a wave moves across the surface of separation of two media there is a change in the direction of propagation, and that the deviation is always in the same sense. The total deviation produced by a prism is therefore brought about in two stages.

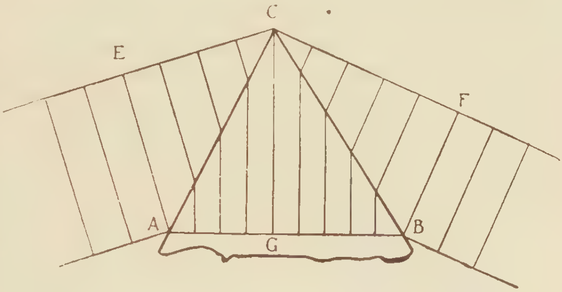


Fig. 175.

The Dispersion of Light. The deviation produced by a prism is really due to the fact that the distance *ECF* in air is greater than the corresponding distance *AGB* in glass. These two distances bear the same relationship to one another as do the velocities in the two media, since they have been traversed in the same time. The total deviation is therefore greater when the refractive index of the prism is increased with respect to that of the surrounding medium. When the deviations for red and green light are compared it is found that they are different. Light waves of different colours have different refractive indices and therefore different velocities of propagation. But since, in air, the velocity of propagation is practically the same for all colours, it follows that in glass it will depend upon the colour of the light. If, therefore, a narrow beam of light containing rays of various colours is allowed to fall on a prism, then the directions of the emergent rays will depend upon the colour of the light in them.

When these rays are allowed to fall upon a white screen the different colours will be observed lying adjacent to one another. It was in this way that **Newton** made the discovery that sunlight consisted of several colours; when he allowed sunlight to pass through a prism he obtained a band of vivid colours. Red light suffered the least deviation; after it came the colours of the rainbow, viz. yellow, green, blue, and violet. Newton was also able to distinguish, between the yellow and red light, another colour, viz. orange; similarly indigo was a colour intermediate between blue and violet. In all, seven colours were distinguished by Newton. We have stipulated that the beam of light investigated must be narrow; this is so because the various colours would otherwise overlap and thereby tend to obliterate one another. The red and violet emergent rays are inclined to one another, the angle between the extreme red and violet rays being called the angle of total dispersion for the given prism, whilst the angle between rays of any other two colours is the angle of partial dispersion. The angle of total dispersion is therefore also equal to the difference between the angles of deviation for red and violet light. This discovery caused the prism to become an important tool in investigating the colours of light. We shall return to this question at a later stage.

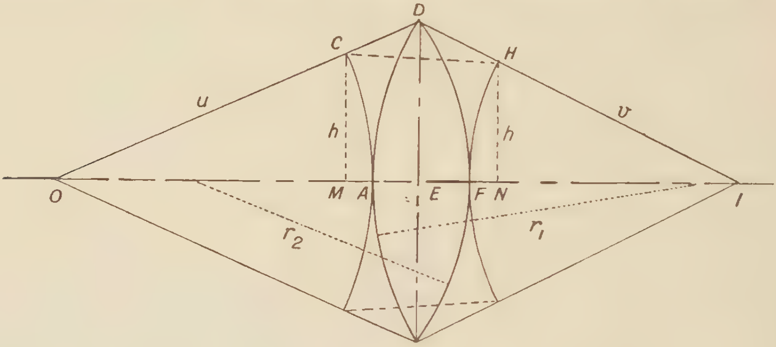


FIG. 176.

Lenses. Lenses, especially those known as burning glasses, were certainly familiar to the ancient Greeks. We are perhaps not very far from the truth when we assume that they were discovered in connexion with spherical glass vessels filled with water and then placed in direct sunlight, so that the sun's rays were brought to a focus at one point.

At all events, burning glasses are repeatedly mentioned by the Greek writers. Lenses play a very important part in most optical instruments which are used to produce an image. The simplest burning glasses, or pocket lenses, are bounded by two spherical surfaces the concave sides of which are turned towards one another. If one only considers lenses whose faces are portions of a spherical surface, they are only to be distinguished from one another by the fact that the radii of curvature of these faces vary from one lens to another. In addition to such lenses, however, there are those designed for special purposes in which the surfaces may be cylindrical, parabolical, or be part of some other surface of revolution.

Let us assume that we have a convex glass lens of the usual type, Fig. 176, r_1 and r_2 being the radii of curvature of its two faces. The line joining the two centres of curvature together, which also passes through the centre or pole of the lens, is known as its principal axis. Any other straight line passing through the centre of the lens is a secondary axis.

A light wave whose source is at O is incident upon the lens, and the problem is to determine how this wave is propagated after passing through the lens. The wave front of the wave when it first touches the lens at A has been drawn in the diagram. Until this wave reaches A it has been travelling without hindrance, so that all points on the front CAK are in phase. From now onwards an ever-increasing portion of the wave will be inside the glass, and when D, a point on the periphery of the lens, has been reached, the distance traversed in the glass will be μ-times smaller than CD. Since the refractive index for glass is about 1·5, we shall assume this value in the present example. Hence the distance AE in glass will be $\frac{2}{3}$ of the distance CD in air. Whilst the central portion of the wave then moves across the distance EF, the edge of the wave front moves along a distance DH, where $DH = \frac{3}{2} EF$. All points on the surface HFJ will be in phase. If now, all these points be on a segment of a spherical surface, then the disturbances will all arrive at the centre of this sphere at the same instant ; in other words, this point will be an image of the luminous point O.

In order to be able to express the position of this point by means of an equation, we have to consider how we have determined the position of the wave after its passage through the lens. We have taken into account the fact that the path in air is μ-times as large as that in glass, both paths being traversed in the same time. Let u be the distance of the object O from the lens, r_1 and r_2 the radii of its faces, and v the distance of the image from the lens, which is unknown. We shall now assume that only those rays are used which lie near to the principal axis, even if, for the sake of clearness in the diagram, more divergent rays have been drawn. By the aid of the lemma on page 394 we can at once write down the magnitudes of all these lengths. We have

$$MA + AF + FN = CH \qquad [\because CH \text{ is approx. parallel to } AF]$$
$$= CD + DH \quad [\text{ since the } \langle CDH \text{ is almost } 180°].$$

But $CD + DH = \mu . AF$

$\therefore \; MA + FN = (\mu - 1)AF.$

With the usual convention with respect to signs [*i.e.* distances measured from the pole against the light are *positive*], r_1 and u are positive, whilst r_2 and v are negative. Now with the same degree of approximation as above, $CM = ED = FH = y$ (say). Then

$$MA = \frac{y^2}{2u}, \; AE = -\frac{y^2}{2r_1}, \; EF = \frac{y^2}{2r_2}, \text{ and } FN = -\frac{y^2}{2v}.$$

$$\therefore MA + FN = \frac{y^2}{2}\left[\frac{1}{u} - \frac{1}{v}\right],$$

and
$$AF = AE + EN = \frac{y^2}{2}\left[\frac{1}{r_2} - \frac{1}{r_1}\right].$$

Hence
$$\frac{1}{v} - \frac{1}{u} = (\mu - 1)\left[\frac{1}{r_1} - \frac{1}{r_2}\right] \quad . \quad . \quad . \quad (191)$$

If, in this expression, we put $\frac{1}{f} = (\mu - 1)\left[\frac{1}{r_1} - \frac{1}{r_2}\right]$, where f corresponds to the value of v for $u = \infty$, we have the usual formula for lenses, viz.

$$\frac{1}{v} - \frac{1}{u} = \frac{1}{f} \quad . \quad . \quad . \quad . \quad . \quad (192)$$

In this expression f is really the focal length of the lens, for parallel rays $[u = \infty]$ will give rise to an image at $v = f$, since $\frac{1}{u} = 0$. For convergent lenses the focus concerned does not lie on the same side of the lens as the object; it lies on the opposite side.

In addition to convergent lenses, there are also those which are **divergent**. Divergent rays, as we have seen, can be caused to converge again by the use of a convex lens, because the central rays travel through a greater distance in glass than do the peripheral ones. In consequence of this, the central portions of the waves are retarded, which so alters the shape of the wave front that the rays are brought to a point. On the other hand, when the peripheral portions of the wave travel a greater distance in glass than do the central ones, then it is the former which are retarded so that the divergence is increased. This is what takes place in a concave or divergent lens.

Chromatic Aberration. In equation (191) the index of refraction μ is a variable depending upon the colour of the particular light used. If, therefore, the source of light is composite in character, all the colours do not give rise to images at the same point. Since the refractive index is greater for blue rays, f must be shortest for them, so that the blue image will be nearest to the lens. The red rays will produce an image at a greater distance from the lens than this. From this it follows that, in the case of a convex lens, the cone of rays leaving the lens will be tinged red on the outside, but after the rays have crossed the principal axis, the peripheral colour changes to blue.

It is therefore impossible to produce an image of a multicoloured point source at a single point with the aid of *one lens only*. The defect in a lens which prevents it from producing a sharp image is known as **chromatic aberration.** In addition, lenses, like spherical mirrors, also suffer from spherical aberration because their spherical shape does not allow all portions of the waves to form an image at the same point. But this defect can be reduced by using only those portions of the waves which are near to the axis, *i.e.* by reducing the aperture of the lens.

Since different glasses have very different dispersive powers, by combining convergent and divergent lenses together, opticians have been able to produce lenses in which, though not for all colours, for two or even three, the images do occur at the same point, so that even the keenest eye cannot detect any error. Nevertheless it is necessary that lenses should be designed for the particular purpose in hand. For images which have to be observed only by our eyes, as with ordinary microscopes, it is necessary for all colours to which our eyes are sensitive, to unite together at one point. But these colours are red, yellow and green. If these colours form images at the same point, then the error produced by the others is only detected when the image is scrutinized very minutely. In photography, however, conditions are different. Since the red and the yellow rays do not, in general, affect a photographic plate, there is very little purpose served by paying any particular attention to them. Photographic lenses should be made so that the especially actinic rays, the violet and ultraviolet, unite to produce a sharp image. During the last few decades opticians have succeeded in producing lenses having a high degree of perfection. In so doing they have helped to bring about the flourishing condition which exists at the present time, not only in physics, but in science as a whole. This happy state of affairs is largely due to the telescopes, microscopes and photographic apparatus which they have designed and perfected. Especial mention must be made of **Abbe** of Jena, who was the first person to introduce exact calculations according to the diffraction theory of light into the science of applied optics, and who, in conjunction with **Zeiss**, made lenses which were far in advance of anything previously attempted. The details of these calculations lie beyond the scope of this present book.

Even in the eyes of men and animals it is the lens which is one of the main components. It is situated behind the pupil and produces on the retina, which is a membrane covering the posterior portion of the eye, an image of external objects. It is true that our eye cannot be extended like a telescope in order to see near and far objects alternately, but since it is a living organ, the curvature of its faces can be altered within limits by means of the ciliary muscles. This change in curvature, of which we are unconscious, takes place until a sharp image has been produced upon the retina of our eye.

Units for Lenses. The power of a lens is measured in **dioptres, convex** lens being reckoned **positive, concave** being **negative.** For purposes of calculating this quantity the focal length is expressed in **metres.** But since a convex lens of large focal length only causes the rays to converge slowly, we have to measure the power of a lens as the reciprocal of its focal length in metres. Thus a lens whose power is $+ 3$ dioptres is a convex lens whose focal length is $\frac{1}{3}$ metre; a power $- 3$ dioptres indicates a concave lens with a focal length numerically equal to that of the above convex lens. We have yet to say, however, what the term focal length implies when used in connexion with a concave lens. If a divergent lens is held in a beam of parallel light, this is caused to diverge as if it came from a point on the object side of the lens. This point is called, in a somewhat loose sense, the focus of the lens. Since it lies on the opposite side of the lens than for convex lenses, we easily see why the focal length of a concave lens is said to be **positive** (although its power is negative).

Defects of Vision are due to the fact that images are not formed sharply on the retina itself, but lie either before or behind it. Such defects can be remedied by the use of suitably chosen lenses.

In short-sighted persons the image is formed in front of the retina, so that they need a divergent lens (concave). Long-sighted people have an eye lens which is not sufficiently strong to concentrate the light rays on the retina; such people need a convergent lens (convex), which helps to increase the power of the eye lens.

3. THE THEORY OF DISPERSION

The chief question for us at present is to account for the fact that light travels through bodies with speeds differing from that with which it travels through the æther alone, and that, even in the same substance, it varies from one colour to another. The elastic theory of light gives us no answer. According to it [*cf.* p. 80] the velocity only depends upon the properties of the medium, and is independent of the way in which the medium is excited. Whether the waves are long or short, the velocity is the same. We shall soon see that different colours denote light waves of different frequencies. We should accordingly expect that all colours would be transmitted with the same velocity. This is certainly so in the free æther, a fact for which we have definite proof.

At the instant when one of Jupiter's satellites [*cf.* Fig. 166, p. 385] emerges from behind that planet, light of all colours is at once sent out. If therefore the blue rays travel more quickly than the red rays, they will be received at the earth before the red rays arrive. The emerging satellite would therefore appear blue to us at first; the other spectral colours would then be added one at a time. When this moon disappeared it would be the longest (red) waves which would bring us the last message, *i.e.* its colour, when last seen, would be red. Since, according to the position of Jupiter, light takes from $\frac{3}{4}$ to 1 hour to reach us from it, a small difference in the velocities of propagation for blue and red rays would make an appreciable difference in the time of arrival. Even with the light from stars, which comes to us across still greater distances, no such differences in velocity for the various colours have ever been detected.

These observations prove that the ultimate cause for the differences which exist between the velocities of propagation is to be sought in the media themselves. But since light waves are exceedingly small, we might expect that an answer to this question can only be given by the molecules themselves. The answering of this question has for a long time occasioned considerable difficulties, and even to-day it has not been finally solved.

Anomalous Dispersion. In the phenomenon known as anomalous dispersion we may perhaps have the clue which will eventually lead to an explanation of dispersion. It is somewhat as follows. For most substances the refractive index increases with the frequency of the light, so that blue rays suffer a greater deviation when they pass through a prism than do the red rays, etc. But there are a few bodies which behave quite differently. In 1870, **Christiansen**, of Copenhagen, made the discovery that a hollow glass prism containing an alcoholic solution of fuchsine produces a spectrum in which the colours are not arranged in their normal order, but as follows—blue, violet, red, yellow. The refractive index for the red and yellow light was greater than for the blue. This can be shown very clearly by means of the following experiment. Using a very narrow slit, an ordinary spectrum is first obtained by means of a glass prism. Let this be *AH*, Fig. 177. Now let a second prism be placed behind the first, its edge being horizontal. If this second prism is made of glass the blue rays will again undergo the greatest deviation and the oblique band of colours *A′H′* is obtained. But when the second prism contains some of the above fuchsine solution, the colours are arranged in the form of two arcs, as shown. These two arcs are separated by an interval representing green light—this is, the light which the solution absorbs. The light waves on either side of this absorption band behave differently. All light waves on the red side of this band are more strongly refracted (in the figure the spectrum lies below *A′H′*), in fact the amount of deviation increases the more closely the wave-length approaches that of the green light which has been absorbed. All light having a higher frequency than that of the green is refracted less strongly (it lies above the normal position *A′H′*). It was **Kundt** (1834–1894) who first pointed out that this phenomenon indicated a connexion between the phenomena of dispersion and the

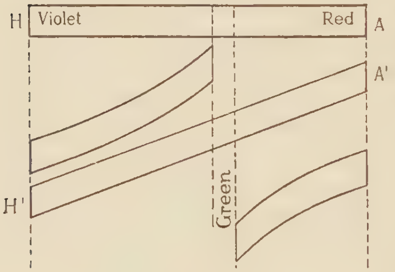


Fig. 177.

characteristic vibrations of some entities inside the various bodies. Bodies absorb from the radiation which is incident upon them just those rays which they themselves can emit : the phenomenon of dispersion therefore proves that these entities can act as resonators for their own radiation. Basing their work on these ideas, various investigators then sought to find a solution to this problem in terms of the elastic theory of light. **Cauchy** and **Ketteler** must certainly be mentioned in this connexion. After the electromagnetic theory had been introduced, attempts were made to obtain a more exact solution in terms of the electrical properties of matter. Guided by probable assumptions concerning this process, an attempt was first made to discover a mathematical formula which would express numerically the relationship between the refractive indices of a medium for various wave-lengths. The success attained was indicated by the extent to which the derived formula agreed with the results of experiment. All investigators assumed a mutual action between the incident waves and the characteristic vibrations of the bodies, but so far, no investigator has succeeded in explaining this reciprocal action, nor in specifying what it is in a body which is capable of executing these vibrations.

4. Optical Instruments as an Aid to Vision

All optical instruments which are used to observe things are designed with the object of increasing the range of vision which our eyes normally possess, for like all our other senses, they can only function between quite definite limits. The sensation of sight can only be stimulated from outside. That lies in the very nature of things, for it is through our eyes that we learn what the objects about us are, and what happens in our vicinity. If these external things are to produce the sensation known as sight, then certain conditions must be fulfilled. For our eyes, as for the other senses, there is a certain minimum value which must be exceeded before they can be stimulated. There is also a second limitation, for not only do we wish to receive a light impression with our eyes, but we also wish to obtain from this impression an idea of the shape, colour, and motion of external objects ; in the simplest case we want to be able to recognize two adjacent points as such. It is only by the different characteristics which we observe in things seen that we recognize them again or distinguish them from one another. Our eyes are not always capable of conveying this information to us. The function of our eyes is to cast upon the retina a continuous picture of things seen. But this ‘ canvas ’ upon which the picture is printed is really a kind of mosaic work, so that it is impossible to see details which are so close together that their images lie closer together than two ‘ tiles in this mosaic pattern ’. The retina consists of a large number of nerve endings all lying very close to one another, and especially of cones and rods. These latter serve to detect a very feeble light impulse. Each cone can only indicate one detail in the object which is being viewed ; two points which lie close together will thus only be perceived as two so long as their images fall upon two different nerve endings. When the experiment is made to see how near together two parallel lines can be and yet be seen separately, we find that the two lines must be at least so far apart that they subtend at the eye an angle equal to about $\frac{3}{4}$ minute [60 minutes of arc = 1 degree].

If we have, for example, two wires near together, then there are two reasons why the angle which they subtend may fall below the above limit ; the two wires may be made to approach one another, or, keeping them at the same distance apart, they may be made to recede from the observer. There are therefore two conditions which determine whether or not we are able to see details in an object: the object may be too remote, or the details may be too small. The function of an optical instrument is therefore to enable us to see what is beyond the power of our unaided eyes. They have to operate in such a way that rays arriving from different points in the object, in spite of the distance of the latter or the fineness of its structure, fall upon different cones of the retina.

Telescopes. A square plate having an edge 1 metre long placed at a distance of 5 km. is observed by our unaided eye as a single point. We are not able to distinguish any details in it. When an astronomical telescope (first invented by **Kepler**) is used to view such an object, Fig. 178, the objective O, or large convex lens, directed towards the object, first produces a real image A_1B_1. In consequence of the fact that the object is a long distance from this lens the image will be formed at its focus. The size of the image is increased by using an objective of longer focal length. We can bring our eye as near as we like to this image and thereby increase the angle between any two rays proceeding from the image to our eye. But our eye cannot perceive things clearly when they are too close to it—the nearest point at which they can be seen distinctly is known as the **near** point. In practice an eye-piece O_1, consisting of a convex lens, is used to magnify the first image, at the same time placing the final image at A_2B_2, so that it can be seen clearly, *i.e.* it is at our near point.

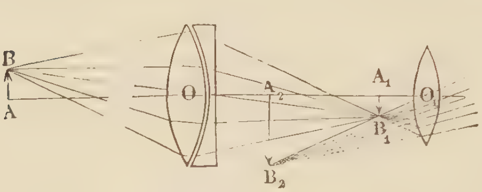


Fig. 178.

The action of a telescope is therefore to bring an object effectively nearer and thereby remove the cause which prevented us from seeing the object clearly with the unaided eye.

The Action of a Microscope. The second fact which limits our vision is that the objects are too small. It might perhaps be thought that in this case the angle subtended at the eye by two points in the object could be increased by diminishing the distance between the eye and object. That is certainly a possibility, but we have to remember that it is necessary for the object to be beyond a certain minimum distance from the eye if it is to be seen distinctly. In a normal eye this is usually about 20 cm. When an object is nearer to the eye than this, the curvature of the incident waves is so great that our eye is not able to cause these waves to converge sufficiently rapidly as to produce a sharply defined image on the retina. If, therefore, an object is smaller than about 0·05 mm., and it is held 20 cm. from the eye, then the angle subtended is less than $\frac{3}{4}$ minute. The details in the object can no longer be discerned, and it is necessary to have resort to a microscope.

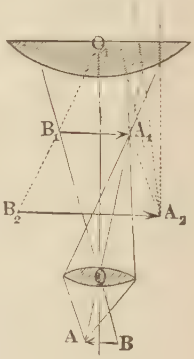


Fig. 179.

Essentially, the microscope consists of an objective O and an eye-piece O_1, Fig. 179. By means of the objective an enlarged image of a small object AB is obtained at A_1B_1, which is again enlarged by the eye-piece O_1 so that the final image is A_2B_2. Since an enlarged image is only obtained when the distance of the image from the lens is greater than that of the object, it follows that, if the dimensions of the microscope are not to make it become cumbersome, the focal length of the object glass must be small. As a rule microscopes are made with a fixed tube length. This limits the distance of the image from the object lens. Hence, with this distance fixed, the magnification will depend upon the distance of the object. Since this is always very close to the focus of the objective, it follows that a big magnification will be obtained by decreasing the focal length of the objective. When the magnification is increased, however, there is always a diminution in the corresponding brightness of the image, since the same amount of light energy is distributed over a larger surface. A second necessity which arises therefore, is that as large as possible a fraction of the energy which the source emits should be utilized in producing the image. In order to achieve this, the object should be brought as near to the lens as is feasible, *i.e.* the focal length of the objective should be small. Different degrees of magnification are therefore obtained, in general, by using various objectives. The desired magnification may also be obtained by the use of another eye-piece.

In connexion with the design of both telescopes and microscopes it is necessary that the objective should produce a sharply defined image, if the final image seen is to be sharp, for no amount of refinement in the eye-piece can compensate for a defect in the objective. In the matter of producing an image, especially if the object is small, the ultimate performance which an instrument can accomplish is limited by a phenomenon known as diffraction. After we have dealt with diffraction processes, mention of this point will have to be made again.

The Examination of Weak Sources of Light. Our eyes are not able to perceive an object when the light energy which they receive from the object is not sufficient to stimulate them. There are, for example, a great number of stars which cannot be seen with the unaided eye. In general, they are so far distant, that all the light rays sent out from different positions in them to our earth are so nearly parallel that the images formed even in the largest telescopes are at their foci. These images are therefore mere luminous points, having no finite size, but just like those obtained when strictly parallel rays are used. But when we direct our unaided eyes towards the star, nothing at all is seen, since the light is too feeble. But if we point the large object glass of a telescope towards it, this lens receives a hundred times the amount of energy that fell upon the pupil of our own eye. All this energy is concentrated in the small image at one point, viz. the focus of the object lens, and can be examined with an eye-piece. The light energy is now sufficient to stimulate the optic nerve of our eye. It is well to notice, however, that we only discern the position of the star and learn that it really does exist, but it is not a true observation in which we recognize the details or shape of the object viewed. Sometimes, however, the light energy which becomes concentrated at a point in such telescopes is still too feeble to be observed, and nothing is gained by merely gazing into the telescope for a long time. There are, however, other means of detection available. If the light from a star of very low magnitude is allowed to fall on the same point of a photographic plate for a whole night long, it is the total energy falling upon the plate which will determine the intensity of the action which takes place. In this way it is possible to record the presence of weak stars which are quite beyond the possibility of ever being observed with the eye.

CHAPTER III

INTERFERENCE AND DIFFRACTION

1. Interference Phenomena

The Colours of Thin Films. The colours of thin films are amongst the first phenomena which are well known to us and which can be attributed to the interference of light waves. If soap bubbles are produced by shaking a soap solution, these bubbles exhibit many beautiful colours. Then layers of turpentine and other substances spread upon the surface of water are also coloured. These phenomena are first mentioned by **Boyle**, whilst **Hooke** (1635–1703) discovered that the colours vary with the thickness of the film. He attributed the colours to the reflexion of rays at the two boundaries of the film. According to him, these reflected rays came together in the eye and there produced the colours.

Newton's Rings. **Newton** investigated the phenomenon more closely, for which purpose he procured thin films of definite and known thickness. He placed a lens having a small curvature upon a flat piece of glass, Fig. 180. The thickness of the air film at any point can be calculated by using equation (188), according to which $d = \varrho^2/2r$, where ϱ is the distance of the point considered from the place where contact is made between the lens and the plate, and r is the radius of curvature of the lens surface. That the thickness of the film does play an important part in this connexion is shown by the fact that the point of contact (central spot) is surrounded by bright and dark circles. When the white light from the sun is used the rings are gorgeously coloured. In general, the colours are the same as are obtained when such light passes through a prism. When, however, monochromatic light is used, such as the light from a bunsen flame into which sodium chloride has been introduced, then the circles are alternately only bright yellow and dark, but the number of rings present is greatly increased. The circles become closer together as their diameters increase and the way in which the diameters vary can be expressed very simply. The radii of successive bright rings, as observed from above, are proportional to the square roots of the odd numbers, whilst the radii of the dark rings intermediate between them are proportional to the square roots of the corresponding intermediate even numbers. Newton deduced from the above that the thickness of the different layers of air where the bright rings were found were proportional to the numbers 1, 3, 5, . . ., whereas, for the dark rings, the thickness was in the ratio of the numbers 2, 4, 6, ... It is very instructive to see how easy it was for Newton to explain these phenomena in terms of his emission theory.

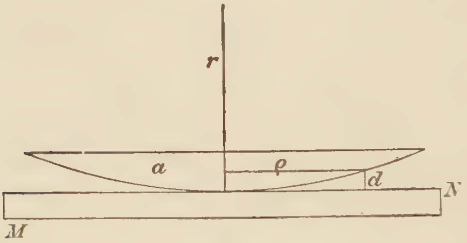


Fig. 180.

He only needed to assume that the travelling particles, which were subject to 'fits', changed from the one type in which they could be reflected into the other when they could be refracted after a definitely short time, and therefore after having traversed a known distance ; these changes took place in a regular manner.

When a particle, being in a 'condition for refraction' at the upper surface of the film, arrives at the lower surface and is just ready to be reflected, then it will be reflected and produce a bright circular ring. But if it should arrive at the second surface in such a condition that it can be refracted, then it passes into the medium below this bounding surface, so that no light would appear above—a dark circle would then be formed. The various bright rings are spaced at such intervals that the thickness of the air layer from one circle to the next increases by the same amount each time.

The explanation of these phenomena **in terms of the wave theory** was only forthcoming at a much later date. We have to thank the English physician **Thomas Young** (1802), and the French physicist **Fresnel** (1805), for this great advance. It helped to bring the wave theory to the notice of the scientific world, which at that time generally accepted the emission theory. The explanation which was now given was much more satisfactory than that of Newton, since it was not necessary to make any additional assumptions.

When light is incident upon a lens-air boundary a portion of the light energy will be reflected, whilst the remainder passes into the new medium, the air film. When this light reaches the flat piece of glass a portion will be reflected upwards so that this wave motion will become superimposed on that which moves upward from the first interface. If these two motions are in phase they will strengthen one another ; if their phases are opposite, then the motions will be destroyed. But since two points in a wave train are out of phase when their distance apart is one half wave-length, it follows that the difference in path for a bright ring and the dark one immediately next to it will be on each occasion equal to one half the wave-length of the incident light.

It must have appeared very remarkable that at the centre of the lens surface, *i.e.* where the lens and glass touch one another, a dark spot was to be seen (when examined by reflected light). Newton was able to account for this very easily. The difference in path for those particles which have been reflected at the first surface at the same time as those at the second surface is equal to zero. Hence all particles capable of being refracted at the upper boundary pass through the lower boundary, since, owing to the fact that the path is zero, they will not have been able to change their state of 'fit'.

The wave theory attributes the presence of this black spot to the fact that the light is reflected at the surface of the lower glass plate, which is a medium more dense than the air through which the light is travelling. This introduces a change of phase [*cf.* p. 83] into the motion, so that the wave moves upward having the same phase as if it had traversed a distance greater by half a wave-length than is actually the case. But since all the waves reflected from the surface of the denser medium are produced under the same conditions, it becomes necessary for us to increase the actual difference in path by half a wave-length for each reflected wave in order to obtain the true phase of the waves. Bright rings will therefore appear when the difference in path amounts to an odd number of half wave-lengths, whilst for the dark rings the difference will be equal to an even number of half wave-lengths. In general, the difference in path is $2d = \frac{\varrho^2}{r}$. If, therefore, λ is equal to the wave-length of the incident light, we have:

$$\left.\begin{aligned}&\text{Bright rings, when } \frac{\varrho^2}{r} = (2n+1)\frac{\lambda}{2}\\&\text{and dark rings, when } \frac{\varrho^2}{r} = 2n\frac{\lambda}{2} = n\lambda\end{aligned}\right\} n = 0, 1, 2, 3 \ldots \quad . \quad (193)$$

It is at once seen that, with the aid of these equations, if r is known, and ϱ measured, it is possible to calculate a whole series of values for the wave-length λ. The radius of curvature of the upper lens is usually made very large in these experiments, in order to increase the distance apart between successive rings. It may be 24 metres or, if all magnitudes are expressed in cm., 2400 cm. If this lens is used to produce the rings with sodium light, then it will be found that the fifth ring ($2n + 1 = 9$) has a radius of 0·8 cm. Hence, from equation (193),

$$\lambda = \frac{2 \times 0{\cdot}64}{9 \times 2400} = 0{\cdot}000059 \text{ cm.}$$

Very careful measurements have shown that the wave-length of sodium light is

$$\lambda = 0{\cdot}0000589 \text{ cm.}$$

Wave-length and Colour. The above phenomenon, due to the interference of light, is the first to provide us with a means of determining the wave-lengths of various lights in absolute measure. Experiment reveals the very important fact that differently coloured waves each have their own definite wave-length. Red light, which consists of the longest light waves, has a wave-length of 0·0000760 cm.; the shortest light waves which our eyes can detect are the violet with a wave-length 0·0000397 cm. The other prismatic colours, yellow, green, and blue, have wave-lengths intermediate between these. Later on we shall become acquainted with other pieces of apparatus by means of which the wave-length can be measured much more accurately than by observations made on Newton's rings. Colour plays a similar part in light, as does the pitch in connexion with sound waves.

The Frequency of Light Waves. Now that we know the wave-length of light and the velocity with which it travels through the æther, we can at once deduce from the equation $c = \nu\lambda$ the number of vibrations executed per second by the vibrating particle, whatever that may be. In the above equation $c = 3 \times 10^{10}$ cm. per sec. Hence for red light, $\nu = 390 \times 10^{12}$ (*i.e.* 390 billions per second), whilst for the shortest violet waves $\nu = 758 \times 10^{12}$. These are enormous numbers, quite beyond our powers of comprehension.

Conditions necessary for Interference to Occur. There are quite a number of methods whereby two light waves can be caused to interfere, *i.e.* alternately reinforce and annul each other. Owing to the shortness of their wave-length, light waves cannot be caused to interfere in the same way as sound waves, although the conditions are fundamentally the same for each. We have to imagine that light waves are sent out in much the same way as sound waves from a vibrating tuning fork. When once a tuning fork has been struck it executes a large number of vibrations; let us assume that it vibrates for 100 seconds before the sound has become so feeble that we cannot hear it. If the frequency of the fork is 500 per second, then the total number of vibrations made by the fork after having been struck once is 50,000. All these follow one another with perfect regularity, if the tuning fork is not disturbed meanwhile. After this the tuning fork must be struck again if it is to vibrate once more. But we cannot suppose that the tuning fork will be excited at just that instant which would enable the second train of waves to be a continuation of the first waves. For example, the maximum condensations in the air would now occur at a distance equal to several wave-lengths behind the condensations in the first wave-train.

In general, waves which are sent out after a fork has been excited once are all at a definite distance from one another, viz. one wave-length; but waves which are sent out after exciting the fork on different occasions are not at regular distances from one another. In order for waves to interfere, it is not sufficient for two waves to arrive on a *single* occasion at the same instant and at the same place, with a phase difference of a half or a whole wave-length. It is true that two such waves would interfere, but since this interference would only take place once, it is impossible for any human being to observe it. In order for interference bands to be seen it is essential that waves shall arrive continuously at one spot with a phase difference of half a wave-length; at an adjacent point the phase difference must be one wave-length; and so on. We can only be certain that these conditions have been complied with if the waves arrive at the point in question from a ‘ tuning fork ’ which has only been excited once, and not when they originate from the same fork excited on different occasions; our certainty is naturally still much less when the waves originate from two quite different forks. Let us now apply these remarks to interference arising from light waves. Two different luminous points, let us say two different parts of a lighted candle or of the sun's surface, correspond to quite different forks. Even waves which are emitted from the same source but which, in the meanwhile, have traversed widely different paths, will no longer have had their origin in the same ‘ strokes of the tuning fork ’. In this connexion we are not able to specify any definite difference in path, since we do not know the time which elapses before a vibrating light-particle comes to rest. On the contrary, we can only conclude that, when two waves still interfere after one has traversed a distance equal to many wave-lengths *more* than the other, then the vibrating element in the light source, between two successive times of excitation, makes a much greater number of vibrations before it comes to rest. We shall, therefore, never expect to produce interference with waves coming from two different sources of light; in fact, it has never been observed. Interference can only be obtained with waves coming from the same *ultimate* source, and such waves are said to be **coherent**, like the waves from the same stroke of a tuning fork.

In the case of Newton's rings coherent waves were obtained by allowing light to fall on the surface of separation of two media (glass and air), where it is partly reflected and partly transmitted. The transmitted portion passes through air to the lower glass plate where it is reflected upwards. It then becomes superimposed upon waves which originated from the source at a later time, and providing these waves were produced during one particular excitation of the ultimate source of radiation, bright or dark fringes will then be produced according to the phase difference at any point.

The vivid colours which manifest themselves when thin films are observed with the aid of reflected or transmitted light are to be explained similarly. A portion of the incident radiation passes through the film and this is reflected at its lower surface. These reflected waves then produce interference with the waves which are being reflected at the upper surface of the film.

When these films are examined in sunlight, some waves, according to the thickness of the film, will have such a phase difference that they destroy one another. These will be absent from the reflected light ; some will be more or less weakened, whilst others, which have not been weakened, will predominate. The reflected light will therefore be coloured differently according to the varying thickness of the film through which the light passes.

In addition to the rays which have been reflected once, there are also others which, having been reflected again at the upper surface, may traverse the film two, or even more, times. Even if the intensity of such rays is not great, they will nevertheless tend to obscure the outer rings.

Concerning the emission of light waves, it is of the greatest importance for us to be able to determine how great the difference in path taken by two waves may be, without losing their capability of producing interference. This problem can be investigated by means of Newton's rings. When the source of light is a sodium flame the number of rings is more than a thousand. For the thousandth bright ring the difference in path between two reinforcing waves will be 1000 λ. We are therefore able to assert quite definitely that even after the lapse of 1000 vibrations, the source from which they originally came is still vibrating and that no new source has been stimulated in the meanwhile. If the rings are kept under observation for a long time, then another excitation will undoubtedly take place, and interference resulting in the formation of the thousandth ring will again occur, when the first member of the wave train has traversed a distance 1000 λ between the two glass surfaces. But this time is so short that no discontinuity will be observed. We have only to remind ourselves of the fact that sodium light, with a frequency of something like 500×10^{12} per second, only requires $0{\cdot}2 \times 10^{-11}$ sec. in order to execute 1000 vibrations. **Fizeau** observed Newton's rings when the path difference had amounted to 30 mm., *i.e.* he was able to count as many as 50,000 rings. With the aid of other pieces of apparatus waves have been caused to interfere when the difference in path has been much greater. **Lummer** and **Gehrke** were finally able to obtain interference when the difference in path was more than a metre, *i.e.* after $2{\cdot}6 \times 10^6$ vibrations.

From this it follows that, whenever the ultimate source which is responsible for the emission of light waves has been excited, this source continues to emit waves for a time which is greater than

$$\frac{2{\cdot}6 \times 10^6}{500 \times 10^{12}} = \frac{1}{200} \times 10^{-6} \text{ sec.}$$

Since gases are also able to emit light radiation, we ought to compare this time with the interval between two molecular collisions [*cf.* p. 208], for we have to assume that, after such a collision, the regular process of vibrating in the molecule is disturbed. At the low pressures to which gases must be subjected in a Geissler tube before light waves can be emitted, the time between two encounters is several hundred times as great as the above period, so that the molecules have sufficient time in which to complete their vibrations.

In all interference phenomena arising in connexion with ‘ thin ’ plates, the whole of the incident wave energy is, as it were, divided into two portions. The one is reflected at once whilst the other passes into the film and is there separated from the first portion for the time being. When these two portions have traversed different distances, we have two waves which are ‘ coherent ’, since they originated in the same source, but with a definite phase difference which we can calculate. Historically these were the first examples of interference to be observed, and for more than a century they were the only ones.

Fresnel's Mirror. In 1816, **Fresnel** showed how coherent waves could be obtained by other means. Light from a narrow slit L, Fig. 181, was allowed to fall upon two mirrors inclined at an angle of nearly 180°, so that they almost lie in the same plane. In this way two images, L_1 and L_2, of the luminous slit are obtained. These images are very close together, the light having been reflected from each mirror appears to come from two (virtual) images. The two trains of waves, one from each mirror, are therefore coherent, since they have their origin in the same ultimate source. These two waves will traverse different distances since they were reflected from different mirrors. Now, whereas Newton with his apparatus as it were ‘ split up ’ the incident light wave, so that every point of the wave supplied a part of the light for the two interfering waves, Fresnel literally cut the wave front in two, and allowed the part reflected from S_1 to interfere with that reflected from S_2. This latter method is very convincing, in so far as the two luminous points L_1 and L_2 and the cones of light from them can be exhibited. At a later date this experiment was so varied that the two separated waves were allowed to form real, but separate, images. In this way it was demonstrated more clearly that the whole process is nothing more than the arrival of two coherent wave trains at the same point.

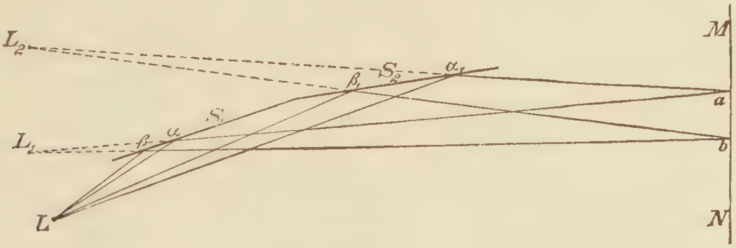


FIG. 181.

Fresnel's Biprism. Two identical vibrating sources of light can be obtained much more simply by allowing light from a point source A, Fig. 182, to fall, partly on the face $\alpha\beta$, and partly on the face $\alpha\gamma$, of a very obtuse angled prism $\alpha\beta\gamma$. The light, having passed through the prism, appears to come from the points A_1 and A_2. But since the two refracted waves originate from one and the same wave they are coherent, and will remain as such. It is just as if the waves had actually originated from the points A_1 and A_2, which sent out coherent waves. On the side of the prism remote from the light, these two waves will therefore be propagated simultaneously and interfere. If the continuous lines in the diagram denote, at some instant, the positions of the waves for which the displacement from the zero position is a maximum in an upward direction, whilst the dotted lines indicate where the displacement is greatest in a downward direction, then a glance at the diagram is sufficient to show that the central line F is cut alternately, first by two continuous lines, and then by two dotted lines. We have now to imagine that the whole picture is moving from left to right with the velocity of light. The point F will therefore be passed by a rapid succession of waves which tend to impart to it a displacement equal to twice that of the above maximum value; this displacement will be directed first upwards and then downwards. The point F will therefore be caused to vibrate vigorously, so that when a white screen is held at F, a bright image appears. On either side of F, however, we see two points D and G at which a continuous line crosses a dotted one. When therefore one wave is endeavouring to urge a particle here upwards, there is another tending to drive it downwards. The resultant effect is that the particle remains at rest, so that darkness prevails at D and G. The points C and H are similarly always bright.

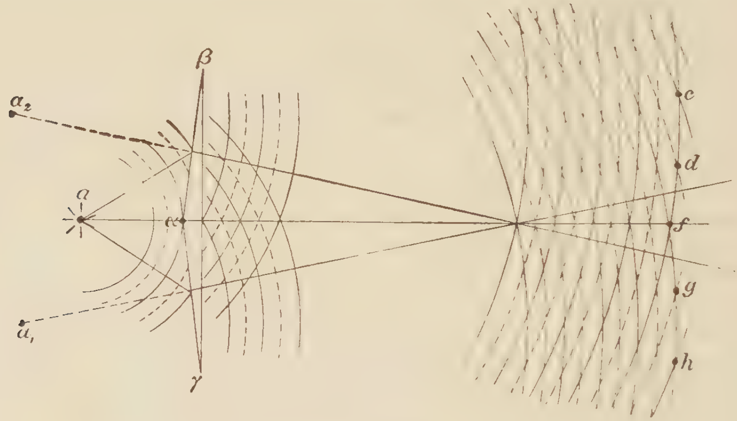


Fig. 182.

If the experiment is made with monochromatic light, *e.g.* sodium light, then the positions of the different fringes can be expressed by the following simple rule. At the first dark band, say D, the difference in path from A_1 and A_2 is exactly half a wave-length; at the succeeding dark bands the differences are 3, 5, 7, . . . half wave-lengths. At the first bright fringe, say C, the difference in path, $A_1C - A_2C$, is a whole wave-length, and at each succeeding bright fringe this difference increases by one wave-length.

In interference phenomena, when light is added to light, the resultant intensity is alternately increased and then reduced to zero. In view of these last-mentioned experiments of Fresnel, Newton's idea of fits of easy reflexion and easy transmission appears to be very improbable. Fresnel himself maintained that these experiments proved that light was a wave motion.

Stationary Waves, produced by interference, were first obtained in 1889 by **O. Wiener**. He allowed light to fall normally on a metal surface AB, Fig. 183. It was assumed by him that the incident waves would interfere with the reflected waves and form stationary waves in exactly the same way as do sound waves under similar conditions. The loops and nodes would be everywhere at the same distance from the reflecting wall and lie in planes parallel to it. The perpendicular distance between his nodes would be half a wave-length.

Much credit is due to Wiener, since he succeeded in detecting these stationary waves. He procured a light sensitive film which was much thinner than half a wave-length of the light used. One end of this film was in contact with the metal A whilst the other end, C, was a few wave-lengths away. He was able to obtain these conditions by exerting more pressure at one edge of the glass which supported the film than at the other. Wherever a loop existed in the film it appeared black on development, whilst the film remained clear at those points where the nodes were situated. Wiener obtained a series of sharp lines on his photographs which were so far apart that they could be seen with the naked eye. The points A_1, A_2, A_3 . . . are the intersections of these lines with the plane of the paper.

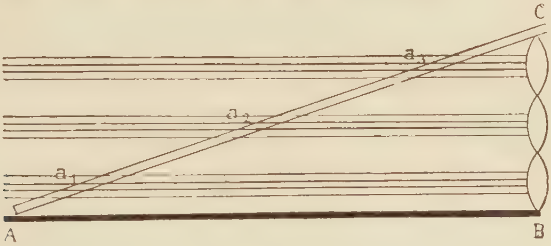


Fig. 183.

2. The Diffraction of Light

Fundamental Phenomena. The Importance of Diffraction. Diffraction phenomena were first observed by the Jesuit priest **P. Grimaldi** in 1661. When he allowed a beam of sunlight to pass through a small aperture in the shutter of a darkened room he noticed that the shadow of a small object held in the beam was broader than could be accounted for by the laws of geometrical optics. Moreover, bright and dark lines appeared on the border of the shadow ; again, geometrical optics did not provide an explanation. When he placed a disc with a small opening in it inside the beam of sunlight, the spot of light obtained on a screen held at some distance from the disc was again greater than that which he deduced by geometrical optics. From these phenomena Grimaldi concluded that he was dealing with a quite new and hitherto unknown property of light ; he called the phenomenon diffraction. On the whole he had at least an inkling of the correct explanation of these phenomena, for, concerning them, he says that when light is here added to light darkness is produced, and compares the propagation of light with that of water waves. But he did not attain to a clear grasp of the circumstances of the problem.

Diffraction phenomena, which are so fascinating when seen, have played a very important part in the history of optics. For the real reasons why Newton could not agree with the assumptions of the wave theory were not that he remained obstinate, in view of the fact that he had developed his own theory, but because, in the first place, it appeared extraordinarily difficult to reconcile the rectilinear propagation of light with the wave theory, and secondly the heavenly bodies moved freely through space, a fact which proved to him that space was absolutely void of material things. It is true that **Huygens** had already attempted to deduce the rectilinear propagation from the wave theory, but his proofs were not convincing. In the case of sound and water waves we can actually see them bending around corners: why should not light waves behave similarly? If Huygens' deductions proved anything at all, they certainly proved it for sound waves; and yet experiment showed that, in spite of this proof, they did not move straight ahead. Diffraction phenomena now showed that light does not always travel in straight lines.

The Explanation of the Rectilinear Propagation of Light. After **Fresnel** had explained interference phenomena in terms of the wave theory, this knowledge enabled him, in 1819, to tackle the problem of the rectilinear propagation of light along the same lines; hitherto this had been the weakest point in the new theory. He commenced with the principle of Huygens, but added to it the knowledge obtained from a study of interference phenomena. This hypothesis was deficient in one respect at least: it did not prove that it was only in a direction perpendicular to the surface *CD*, Fig. 38, page 81, which is the envelope of all the secondary waves, that most of the light energy would be propagated. Fresnel now showed, what we have already explained on page 82, that light could not travel in any of the other possible directions because the secondary waves, proceeding from the different points on the surfaces already reached by the waves, mutually destroy one another. When the validity of these ideas was confirmed experimentally, the death-knell of the emission theory was finally sounded. At the present time we have several such proofs. Let us suppose that alternate bright and dark rings are drawn upon a plate *PP*, Fig. 184, in such a manner that they correspond to Fresnel zones for light of definite wave-length λ, and for the point O. The central disc of such a zone plate is bright, the first ring dark, whilst the 2, 4, 6, . . . rings are bright, etc. If the plate is illuminated with light of wave-length λ, the light which reaches O will always be in the same phase as the light which arrives there from the centre of the disc, because all the zones which would supply other vibrations have been screened off. A bright luminous point will therefore be formed at O: it is just as if O were the focus of a convex lens. The increase in luminosity at O is merely due to the fact that some zones, by being blackened, are prevented from sending waves towards O.

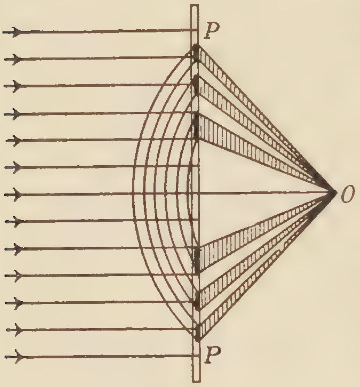


FIG. 184.

Fundamental Ideas concerning Diffraction. From the above remarks one can at once obtain all the fundamental notions which are necessary in order to be able to understand diffraction phenomena. If the effect of any one zone at the point O is to be eliminated, then the two neighbouring zones must cooperate in so far as their effects at this point are concerned. Should these two zones be lacking, or not be able to take effect, then the effects due to the remaining zones will be changed. Where the missing zone should have operated to produce darkness, darkness will no longer occur.

All diffraction phenomena are really a proof of the validity of Fresnel's assumptions. Examples of diffraction are now so numerous and various that we shall only be able to discuss the most important of them; we shall select those from which a real understanding of the subject can be gained.

In order not to complicate matters unnecessarily, we shall always assume that the source of light is at an infinite distance. The wave surface which we have to discuss is then a portion of a plane. Such plane waves are very easily obtained. We can either use sunlight, or else place another source of light at the focus of a convex lens, where the emergent rays are strictly parallel to one another. Moreover, we shall at first assume that the light is monochromatic, *i.e.* the wave-length is quite definite. If that is not so, then the problem has to be solved separately for each wave-length, the combined effect at every point afterwards being obtained by a process of summation. In addition to the calculated variation in intensity from point to point, there would also be a marked variation in colour as a result of this superposition of rings of different wave-length.

Diffraction at a Circular Aperture. Our first problem concerns the passage of light through a small circular aperture. It must not be imagined that it suffices to allow sunlight to fall upon a small round opening, for the effects produced by waves coming from different points in the sun overlap and obliterate one another. It is first necessary to isolate a single point source from the whole of the radiation coming from the sun, as **Grimaldi** had done previously. He used the light which passed through a small opening in the shutter of his room. Behind this were placed in succession two other small openings so that eventually only light from a very small portion of the sun's disc passed through. Let us now suppose that such light falls upon a small circular aperture *PQ*, Fig. 185, similar to that made by pricking some tinfoil with a needle. In the diagram this opening has been enlarged considerably so that the wave-length of the light can be represented on a convenient scale.

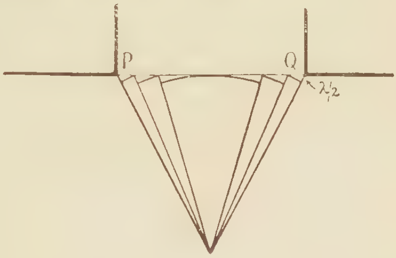


FIG. 185.

Let us first calculate the intensity of light at O, a point on the normal through the centre of the opening. For this purpose it is only necessary for us to proceed in exactly the same way as we have already done in connexion with the principle of Huygens and Fresnel. All points in the plane PQ are arranged in groups by constructing the Fresnel zones. If a is the distance of O from the centre of the aperture PQ, with ~~P~~ O as centre, we draw circles of radii a, $a + \frac{\lambda}{2}$, $a + \frac{2\lambda}{2}$, $a + \frac{3\lambda}{2}$, . . . Each zone is at a greater distance from O than its predecessor by an amount $\frac{\lambda}{2}$. On account of the smallness of the aperture only a few zones will be found in it. If the number of zones is even, then the effect at O will be darkness, since their effects destroy one another; if the number is odd, then the effects of all zones except the first will be effaced.

The nearer O lies to PQ the more numerous will the zones be. If, for example, O is at an infinite distance away, there will only be one zone inside PQ. When therefore the distance of O from the aperture alters, this latter will contain first an even number, and then an odd number of zones. The point O is therefore bright and dark alternately.

At points a little to the right of O, and also even at points in the geometrical shadow of the screen, the intensity will be different from that at O; first there will be a brightness, say, and then darkness. Since the figure which we have drawn is really a section through circular surfaces and a cone, the actual pattern will consist of circular fringes around O. But since those points at which the brightness is a maximum are not immediately adjacent to those where it is a minimum, it follows that there will be a gradual transition from light to dark.

Diffraction Bands due to a Circular Disc. The next problem is to examine the diffraction bands produced when a circular disc, Fig. 186, is placed in front of a luminous source. We can again draw the zones corresponding to a point O on the axis of the figure. The central zones are now absent entirely; there remain only the zones lying beyond the edge of the disc, and they will of course extend to infinity. Now according to the principle of Huygens, already discussed in detail, the effect of each zone is balanced by the effects due to those of the half zones which are adjacent to it. The resultant effect is therefore due to one-half of the peripheral zone of the disc. Since the point O is equally distant from all points on the edge of the disc, a bright luminous point will be seen at O. This image will be surrounded by a system of dark and bright fringes.

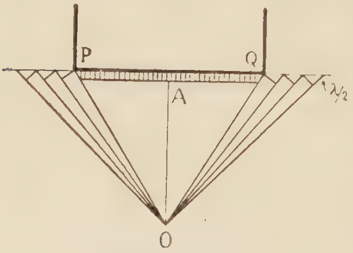


Fig. 186.

Let us now assume that Figs. 185 and 186 are really sections, one of a long narrow slit of width PQ, the other of a long wire of diameter PQ. The remarks which we have already made will still be applicable to these new problems, but with this difference—the fringes will no longer be circular but will run parallel to the length of the slit or wire. In the middle and extending the whole length of the wire we shall find a bright line.

The Measurement of Wave-lengths by means of Diffraction Bands. The arrangement just described can be used to obtain a measure of the wave-length of light. Let $AB = a$ be the known diameter of a needle, Fig. 187, placed at a distance h from a screen. Let $ED = d$ be the distance of the first dark fringe; then it is known that $EB - EA = \frac{\lambda}{2}$. Moreover, it can easily be shown that this difference can be expressed by

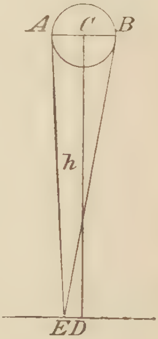


Fig. 187.

$$EB - EA = \frac{ad}{h}, \text{ so that } \lambda = \frac{2nd}{h}.$$

The experiment can be repeated using light of different wave-lengths, which can then be measured with a fair degree of accuracy because, by making the distance h large, ED can be made as big as is desired. The distances which have to be measured are then all considerable, and therefore permit exact measurements to be made.

Diffraction Gratings. The diffraction phenomena produced when the source is at a very large distance away, *i.e.* parallel rays are used, are very important. These can be quite well observed by allowing the rays to fall upon a convex lens of considerable focal length (as in an astronomical telescope). A diffraction pattern is produced at the focus of the lens, and this can be observed by means of a second convex lens of shorter focal length. Perhaps still more important are the diffraction patterns produced by means of a large number of lines placed at equal distances apart; we refer to the so-called gratings. Such gratings are produced by ruling, with the aid of a suitable machine, upon a piece of plate-glass or carefully polished metal (speculum), a large number of lines at equal distances apart. The lines are ruled by means of a diamond. The most famous of these gratings are those produced by the American physicist **Rowland.** They may have as many as 40,000 lines to the inch, with a total of 100,000 lines to each grating. In the case of metal gratings the diffraction phenomena can only be obtained with reflected light. Sometimes the lines are ruled upon a concave or cylindrical surface, when the diffracted image is formed at a sharp point or a narrow line. The great advantage of these concave gratings lies in the fact that it is not necessary for the light to pass through glass, so that it is possible to investigate those wave-lengths which are so short that they are absorbed by glass.

Let AH, Fig. 188, be the section of a plane grating, which is perpendicular to the plane of the paper. Let a be the width of the clear space between two consecutive rulings, whilst b is the width of each ruling. The distance $(a + b)$ is called the **grating constant.** We will suppose that a beam of parallel light is incident upon the grating. The secondary wavelets which are developed in the clear spaces will spread out in all directions. We will consider those wavelets which make a certain angle, θ, with the normal to the grating. If these wavelets encounter a convex lens, L, placed parallel to the grating, they will be brought to a focus at P say, so that the illumination at P will be the resultant effect due to all the wavelets we have considered. If the disturbances reinforce one another at P, a bright image will be produced there, otherwise it will remain dark. In order to determine the difference in phase of the waves which arrive at this point, draw BF perpendicular to the rays. Since the optical path is the same from B to P as from F to P, it follows that any phase difference will be introduced in the distance DF.

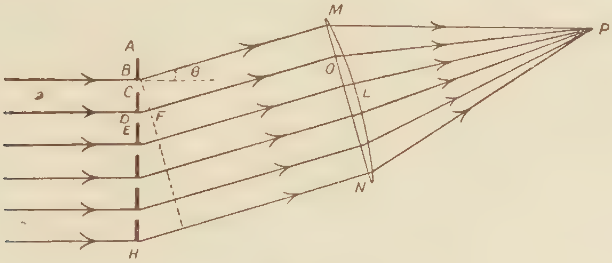


FIG. 188.

Now the difference in path between the two rays BMP and DOP is DF, or $(a + b)\sin\theta$. The point P will therefore be bright if

$$(a + b)\sin\theta = n\lambda \quad . \quad . \quad . \quad . \quad (194)$$

where n is any integer, including zero. On the other hand darkness will prevail at P if

$$(a + b) \sin \theta = (n + \tfrac{1}{2}) \lambda \quad . \quad . \quad . \quad (195)$$

If instead of monochromatic light, rays of different wave-lengths are used, each new wave-length will give rise to a position of maximum brightness different from that just determined, since $\sin \theta$ changes with λ. The various colours will be observed in adjacent positions, as with a prism. The spectrum produced by a diffraction grating, however, is to be distinguished from that due to a prism by the fact that with gratings $\sin \theta$ is directly proportional to the wave-length, whereas with prisms the dispersion is different according to the material from which they have been made, and the violet is deviated more than the red.

Since the grating constant is known from the dividing engine used to rule the grating, and the angle can be read off from a graduated circle very easily, gratings furnish a means whereby wave-lengths can be measured with precision.

From the equation (194) it is seen that the angle θ becomes smaller when light of shorter wave-length is used. If, therefore, **X-rays** having a wave-length one thousand times smaller than that of light are examined by a grating, the positions of maximum and minimum intensity are so close together that they can no longer be distinguished. But the smaller the value of the grating constant a becomes, the greater will θ be for the same wave-length λ. We see therefore that determination of the wave-lengths of X-rays only becomes a possibility, if gratings finer than those used for light can be found, unless these are used at almost grazing incidence. Crystals have such a fine and regular structure that they act as suitable gratings for the production of X-ray spectra.

The Resolving Power of Optical Instruments. Let us now return to consider the images produced by optical instruments, and the part played by diffraction in diminishing their definition. We have already seen that when light passes through a narrow opening, diffraction bands appear, because at any point in the image the distances from different points on the periphery of this diaphragm or ‘ stop ’ are unequal. The first dark ring will be formed where the path difference is equal to one half wave-length. The diameter of this circle will be greater, the further the image is from the stop, and the smaller the stop becomes. But inside this dark ring a sharply defined bright point will certainly not be found, for at small distances from the point at which the image is formed according to the laws of geometrical optics, the phase will gradually increase from zero at this point until it is equal to $\lambda/2$ at points on the first dark circle. The image will therefore not have a definite boundary, *i.e.* it will not be limited by a sudden passage from brightness to darkness, but there will be a small bright disc completely surrounded by a set of circular diffraction bands. The greater the opening in the stop, the smaller do these diffraction rings become, and that is the reason why we do not usually see them. They lie so closely together that our eye fails to resolve them. If we wish to observe them, they must be magnified by decreasing the size of the stop.

Such stops are to be found in all optical instruments, for even the mounting in which a lens is fixed constitutes a stop, and therefore causes diffraction rings to be formed. In a simple **astronomical telescope,** the mounting which holds the objective in position prevents the outermost Fresnel zones from assisting in the formation of the images. If we therefore assume that we had an otherwise perfect objective which would focus parallel rays at a point, yet, on account of diffraction, rays coming from some distant star would not produce a point image—instead there would be a small luminous disc at the focus, surrounded by a system of diffraction rings, the diameters of which increase with the focal length of the objective. On a former occasion we found it advantageous to increase the focal length as much as possible ; we now see that this entails a corresponding disadvantage. It is still possible, however, to diminish the size of these rings by increasing the diameter of the stop which, in our case, means that the diameter of the objective must be enlarged. This is the real reason which compels astronomers to use objectives of large diameter and focal length. The efficiency of a telescope is ultimately determined by these diffraction phenomena. If, for example, two stars are so close together that their diffraction discs partly overlap one another, however much the magnification may be increased, two separate images will never be obtained. The limit to which the resolving power of a telescope can be pushed is therefore reached when the diffraction images of two neighbouring stars overlap.

With regard to microscopes the conditions are somewhat more complicated, since one usually wishes to observe the details in objects, which are very small. Moreover, these objects are not luminous points but are illuminated by transmitted light. When this light encounters the fibres, small grains, etc., in the object, it is affected in much the same way as light passing through a grating. But there is one very great difference. In the grating the diffracting centres are at regular intervals apart, whereas in an object these centres are irregular in length, distance apart, and orientation.

The waves diffracted by these centres will therefore not be propagated in one direction, but will spread outwards in all directions. The theory which was developed very thoroughly by **Abbe,** in connexion with this problem, indicates that the image will be more perfect when the number of diffracted rays forming the image is increased. This means that with microscopes as great a portion as possible of the spherical waves sent out shall be utilized in producing the image. Now the angle between the central and extreme rays, this latter being the ray proceeding from the centre of the object to the periphery of the objective, is called the **aperture** of the objective. The above remarks indicate that it is desirable to use an aperture as large as possible. The shape and size of microscope objectives are determined by these considerations.

Resumé. Even apart from the considerable rôle played by diffraction phenomena in connexion with the design of optical instruments, they form a very important section of our study, since the ultimate proof, that light consists of a wave motion, is obtained from the effects due to diffraction and interference. When Fresnel succeeded in accounting for the rectilinear propagation of light in terms of the wave theory, and substantiated his deduction by means of experiment, the problem was really decided, even before Foucault, by determining the velocity of light in water, had demonstrated that Newton's theory of refraction was impossible.

It is through the phenomena of interference and diffraction that our knowledge of the wave-length of light and the period of a single vibration has been acquired. This possibility of measuring the wave-lengths of various light waves with such precision has been, and still is, of the greatest service to those who conduct researches into the nature of the material world. We refer to the subject of spectrum analysis, concerning which a short discussion follows.

Spectrum Analysis. The term spectrum analysis refers to the investigation of bodies by means of their light spectra. It has been shown that all light emitted by the sun and other luminous bodies is not of a simple nature, *i.e.* it does not consist of light of one wave-length, but is a mixture of very different wave-lengths. Even sound consists very frequently of a mixture of various tones, *i.e.* notes of a different pitch. Each individual wire in a piano sends out a great number of different notes, the fundamental and the overtones, which can be detected by the ear of a musician, without any external aid. Our eyes are not able to behave in a similar capacity towards light waves, for various coloured lights produce in them quite a new colour, in which the components are no longer recognizable. Green and red light, for example, produce the sensation of white. But even yellow and blue produce this same sensation of white, and no eye can distinguish whether or not the white light consists of red and green, yellow and blue, or some other mixture of colours. Our eye alone is not able to analyse light into its constituents. The components which may be present in any colour must first be separated if our eye is to see them individually. When the colours have been thus separated, we have a spectrum. We can produce a spectrum by allowing the light under investigation to pass through a glass prism (Fig. 189). For this purpose a narrow slit *G* is illuminated by the light, which then passes through a convex lens *L*, so that an image is produced at *B*. When a prism is inserted in the paths of the rays, the image is formed below *B*. The magnitude of this shift depends upon the refractive index of the glass, which varies from one colour to another. An image of the slit will therefore be formed for each different colour present in the incident light. When we speak of bright or dark lines in a spectrum, these lines are really nothing more than images of a very narrow slit. Each bright line is an image due to one particular colour, whilst each dark line represents a colour which is lacking in the original light. Narrow slits have therefore to be used, otherwise the images produced by rays of approximately the same wave-length will partly overlap. The distances between these lines, and the order in which they occur, depend upon the material of the prism.

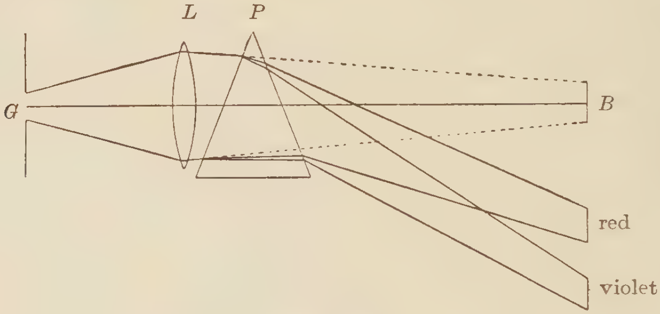


Fig. 189.

Diffraction spectra are produced in an analogous manner, except that a grating is used instead of a prism. By means of the grating the light waves are diffracted in a manner already described. Since, according to equation (194), the angle of diffraction depends upon the wave-length, waves corresponding to the various colours will be diffracted differently and so produce a spectrum.

In both instances we really obtain images of a slit lying at points not on the primary beam. But whereas with prisms the dispersion cannot be represented by any simple expression, with diffraction spectra the dispersion is regular and is related to the wave-length by the simple formula (194).

We have already seen [*cf.* p. 354] that bodies at high temperatures, or under other conditions of excitation, emit light waves; the spectrum produced by these waves is known as an **emission spectrum.** The emission spectra associated with solid bodies are continuous, whereas for gases, the radiation comprises waves having sharply defined lengths.

Gases, however, just like bodies which appear to us to be transparent, absorb some wave-lengths from the light which passes through them; according to Kirchhoff's law [*cf.* p. 334] these wave-lengths correspond to those rays which the gas would emit at the same temperature. When therefore light from a white hot body passes through a gas and its spectrum is afterwards examined, it is found that the lines which are missing are just those which the gas would send out. A gas can therefore be recognized from the light waves it absorbs from white radiation. This is known as an **absorption spectrum.**

Celestial Matter. Only one of the many important services which spectrum analysis has rendered to astronomy can be mentioned here. When, in 1812, **Fraunhofer,** at Munich, produced a solar spectrum in the manner described above, he obtained a continuous spectrum crossed by a series of dark lines. He was able to count 700. Nowadays we know that there are more than 60,000 such lines, some of which are in the visible spectrum, whilst the remainder are to be found in the ultra-violet and infra-red regions. Fifty years elapsed (1863) before the significance of these lines was discovered by **Kirchhoff.** They are due to elements present, either in the chromosphere of the sun, or the atmosphere of our earth, which absorb light of some particular wave-length when it is incident upon them. After the lines due to absorption in our own atmosphere had been discovered, those due to the sun's atmosphere really act as messengers bringing us information about the elements present in the sun's atmosphere, and thus in the sun itself, in so far as there is an exchange of material between the sun and its chromosphere.

In this manner we have learned that the sun and stars consist of the same elements as are found here. Since we are able to measure light waves with extraordinary precision, this information, which has been deduced from many such exact measurements, is beyond all doubt.

Another advantage accruing from the investigation of matter by means of its light spectra lies in the fact that only a very minute quantity of substance is required. For example, even such a small quantity as one forty-millionth part of a milligramme of sodium can be detected in this way. It has been calculated that 7×10^{-14} mgrm. of hydrogen are sufficient to be observed in a Geissler tube. But even this quantity contains more than 40 million atoms. With this one must compare radioactive changes, which have supplied us with information about a single atom. By means of spectroscopy a whole series of new elements have been revealed. Even in their very early work, **Bunsen** and **Kirchhoff** discovered two new elements—cæsium and rubidium. A little later thallium, indium, and gallium were found, and then, in 1895–1898, **Ramsay** detected the presence of the whole series of inert gases in our atmosphere.

The study of spectroscopy received a quite unsuspected development, when modern researches on the atom succeeded in revealing the relations between the atoms and the X-rays which they emit. All the elements are characterized by means of their X-ray spectra with the same certainty as by their light spectra, and this statement is independent of every hypothesis concerning the mechanism of the emission of the radiations and also of the validity of Bohr's atomic theory. Moreover, X-ray analysis possesses a considerable advantage over ordinary methods of spectroscopy. In either method only a very small quantity of material is necessary, but it is not necessary for the production of X-ray spectra that the substance should be in a pure state or gaseous ; it is sufficient if the element is available in an ore or crystal along with other elements. If this element forms part of the anticathode of an X-ray tube, it can be recognized from the X-rays which are emitted. The importance of this method of analysis is enhanced by the fact that it enables us to determine the atomic numbers of the elements, and hence approximately their atomic weights and other properties. By this method **hafnium** (atomic number 72) was discovered in Bohr's laboratory in 1923, by Coster and Hevesy. Since then, as already mentioned on page 369, several other elements have been discovered.

CHAPTER IV

THE POLARIZATION OF LIGHT

1. Polarization by Reflexion

Fundamental Phenomena. A very remarkable property of light was discovered by **Malus** at Paris in 1808. When he was examining a beam of sunlight through a crystal of calcspar, after it had passed through a window in the Palais de Luxembourg, he noticed that one of the two refracted rays normally present disappeared, when the crystal was held in a certain position. This position was obtained by rotating the crystal around an axis parallel to the incident beam. He also investigated the light reflected from glass and water surfaces, and found that it was characterized by a certain two-sidedness, *i.e.* it does not present the same aspects on all sides. Different directions in the plane perpendicular to the beam were no longer equivalent. The idea of polarization has already been introduced whilst discussing waves in general, and we have to remind ourselves of the remarks which were then made. Malus soon discovered the exact rule concerning the conditions under which polarized light can be obtained. When light was caused to fall on a glass plate at an angle of incidence equal to 54° or 55°, the reflected light exhibited the following properties. Let AB, Fig. 190, be the direction of the incident ray which encounters a plate of black glass. BE is the reflected ray falling upon a second and similar glass plate GH parallel to CD. The reflected ray is EF. This reflected ray is quite easily observed. But if the second plate is rotated around an axis parallel to the direction BE of the light incident upon it, the reflected ray is no longer in the plane of the paper, and it gradually becomes less intense. When the rotation has become equal to 90° the intensity of the reflected beam is practically zero. If the rotation is continued the intensity gradually increases until the rotation has amounted to 180° when the intensity is again a maximum. It again becomes a minimum when the rotation is 270°. If the ray BE reflected from the first mirror were to behave in the same manner with respect to all planes passing through BE, then no difference would be observed during the rotation of the upper mirror, so long as the angle of incidence remained the same. From this fact **Young**, the physician whom we have already mentioned, concluded in 1817 that, in light, we were dealing with transverse waves.

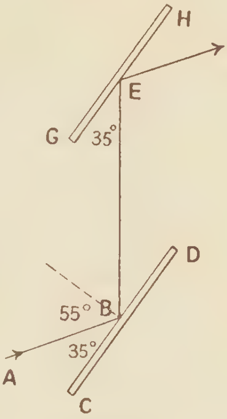


Fig. 190.

When the light is incident at some angle other than about 55° there is still a variation in the intensity of the light as the upper mirror rotates, but it is never reduced to zero. The angle of incidence necessary for the complete extinction of the light varies according to the particular glass used—in fact it also changes with the colour of the incident light. The simple law which defines the angle of incidence necessary for the intensity of the light to become zero, was discovered by **Brewster.** It states that the reflected ray will be completely polarized when it is perpendicular to the refracted ray. Let AB, Fig. 191, be the incident, BC the reflected, and BD the refracted rays. The reflected ray is polarized completely when BC and BD are mutually perpendicular. If α is the angle of incidence, and β the angle of refraction, then $\sin \alpha = \mu \sin \beta$, where μ is the refractive index of the medium. But, in this particular instance, $\sin \beta = \cos \alpha$, since $\alpha + \beta = 90°$. We therefore have

$$\tan \alpha = \mu \quad . \quad . \quad (195)$$

Since the value of μ depends upon the colour of the light, it follows that all colours will not be extinguished at the same time, *i.e.* for one particular value of α. This is why, in the apparatus described above, and using white light, the intensity never actually becomes zero, and why the polarizing angle cannot be stated definitely, without giving the corresponding wave-length at the same time.

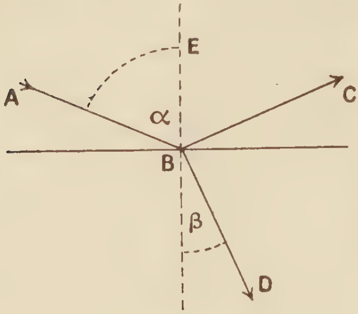


FIG. 191.

For experiments on polarized light two pieces of apparatus are necessary; the first is to produce the polarized light, the second to detect it. In the above experiment two glass plates were used. The first served to render the reflected light polarized. But this change could not be observed since our eyes are not able to detect polarized light directly. The second plate revealed this property of the light. The first is the polarizer, the second the analyser. In all cases the polarizer and analyser may be identical in their construction, but the duties which they perform are quite different, for the first plate receives rays of ordinary light, whilst the second only receives light which has been polarized by the first plate. We therefore have to imagine that in ordinary light the vibrations take place in all directions. But for a fixed angle of incidence, the plate can only reflect those light waves which are vibrating in one particular plane. Let us assume, since we cannot at present decide definitely, that these vibrations take place in a plane perpendicular to that containing the incident ray and the normal to the surface. Then all possible vibrations for the incident light can be resolved into two directions, one of which is in the plane of incidence and the other perpendicular to it. Now that component of the light which is vibrating perpendicularly to the plane of incidence will be reflected, whereas the other component will not. The light which is incident upon the second mirror will therefore only be vibrating in one plane. If this plane is such that reflexions from the second plate can take place, then the polarized light will be reflected. But if it so happens that the plane is perpendicular to this direction, then it will receive no light which it is able to reflect.

2. POLARIZATION BY REFRACTION

Since it is only light vibrating in one plane that is reflected, we have to assume that the remaining portion of the light, *i.e.* the part which is refracted, must at least be partly polarized, since most of the light vibrating in one direction will have been reflected back into the first medium.

This deduction has also been confirmed. The refracted light which passes through the first glass plate is also polarized, but only in part; the vibrations in the two component directions are not equally strong, but each always differs from zero.

If this refracted light is allowed to pass through a second glass plate, then only a portion of the light is reflected, so that the other component assumes a greater relative intensity in the refracted beam. After passing through about twenty such plates, the component which vibrates in one direction has been removed entirely by reflexion, so that the light remaining in the refracted beam is practically only vibrating in one plane. A pile of glass plates, Fig. 192, can therefore be used to produce polarized light; both the reflected and refracted beams can be utilized for this purpose, but the two beams, *CD* and *BF*, are polarized in planes at right angles to each other. If they fall on similarly orientated mirrors, the intensity of the one after reflexion is a maximum, when the other is a minimum.

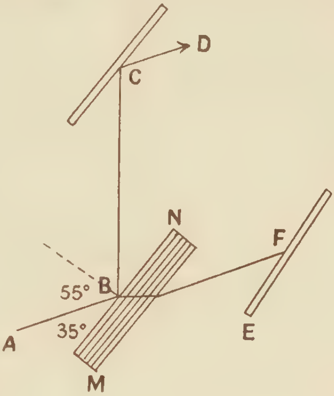


Fig. 192.

Interference of Polarized Light. It is very important that we should be able to determine the conditions under which polarized light can produce interference. By the year 1816 **Fresnel** and **Arago** had shown that two light rays, polarized in planes perpendicular to one another, could not interfere under any circumstances, even if they did originate in the same source. This fact is in perfect accord with the theory of polarized light. For two vibratory motions taking place in planes which are mutually perpendicular can under no circumstances destroy one another. They will always produce a vibration in a plane at 45° to either of the original planes, if the two disturbances pass through their zero positions simultaneously, or else they will produce a motion in an elliptical orbit when they do not so vibrate as to fulfil the condition just mentioned. In particular limiting cases the ellipse becomes a circle.

3. DOUBLE REFRACTION

The phenomenon of double refraction was investigated by **Huygens,** who, with remarkable ingenuity, offered an explanation of it which is essentially correct, although at first sight the phenomena appear to be very complicated. This phenomenon can be observed with all crystals except those belonging to the first, or cubic system, but Iceland spar (calcite, $CaCO_3$) exhibits it most strikingly. It was with this substance that **Erasmus Bartholinus** (1625–1698), a native of Denmark, discovered this remarkable phenomenon. In the remarks which follow we shall confine our attention to this one substance. Calcspar belongs to the hexagonal system [*cf.* p. 138]. It possesses one direction which is distinguished from all others by certain properties; it is called the **optic axis.** If a plane is drawn perpendicular to this axis, then all directions in this plane are absolutely equivalent. A crystal of calcite is bounded by six rhombohedral faces ; they are known as cleavage faces, since it is in directions parallel to them that the crystal can be split most easily. In order to obtain a mental picture of the form assumed by a crystal of calcite, it is best to commence with that of a cube. Let us suppose that a cube, Fig. 193, has one of its corners *B′* resting on a table, and that pressure is applied at the opposite corner *D*. If we now imagine that the cube yields under the influence of this pressure but in such a way that its faces remain planes, then a rhombohedron is obtained ; this is the shape of a calcite crystal. The direction *DB′*, along which the external pressure is applied, and which becomes the shortest diagonal joining two opposite solid angles, is the optic axis for calcite. Every plane in the crystal which is parallel to this optic axis is called a **principal plane.**

Let us assume that we have a crystal with its natural (cleavage) faces, and that light, let us say sunlight which has passed through a small hole in a blind, traverses the crystal and is afterwards received upon a screen. Two bright images will be observed. If the crystal is rotated in the direction of the incident light as axis, then one of the two images remains stationary, whilst the other moves in the same direction as that in which the crystal is rotated. Huygens investigated these rays and soon discovered that the ordinary law of refraction is valid for the ray producing the stationary image. This is known as the ordinary ray, *O*, whilst the other is the extraordinary ray, *E*.

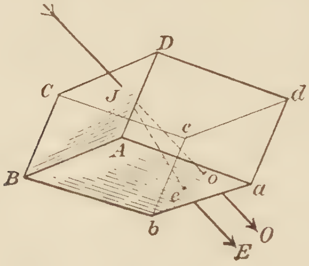


Fig. 193.

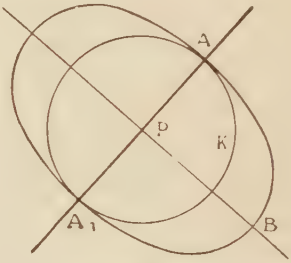


Fig. 194.

Huygens at once formed the idea that these two rays must travel with different velocities in the crystal, and from measurements of the refractive indices he arrived at the following conception. If *P*, Fig. 194, is the point in a crystal of calcite at which a light wave impinges at any instant, then at any other later instant the ordinary wave is the surface of a sphere drawn around *P* as centre. In other words, the disturbance caused by the ordinary wave travels with the same velocity in all directions.

But the extraordinary wave is propagated with a velocity which depends upon the particular direction considered. Now since calcite possesses an axis around which its properties are identical in all directions, the velocities which may be determined for the extraordinary ray must remain the same when the whole is rotated about the optic axis. The wave-surface for the extraordinary ray must therefore be a surface of revolution, and Huygens at once made the assumption that it was a spheroid or ellipsoid of revolution [the simplest possible assumption that it is a sphere is, of course, excluded for the extraordinary ray]. Since the refractive index for the extraordinary ray is less than that for the ordinary ray in calcite, it follows that the velocity for the extraordinary wave must be the greater. The separation between the two rays becomes less as the direction of the incident beam approaches that of the optic axis ; when these two directions coincide the two rays coalesce. Huygens therefore made the additional conclusion that the two rays travelled along the optic axis with identical velocities, since they do not become separated from one another. The sphere must therefore touch the ellipsoid at the extremities of its minor axis. Huygens was able to confirm this idea by cutting a crystal of calcite so that one of its faces was perpendicular to the optic axis. When a ray of light fell normally on this face, and therefore in the direction of the axis AA_1, only one ray was obtained ; for any other direction of incidence two rays were always present. The axis of rotation for the ellipsoid must therefore coincide with the optic axis. When, therefore, the wave-front for the ordinary ray had spread from P so that it coincided with the surface of the sphere K, that of the extraordinary ray had reached the surface of the ellipsoid which touches the sphere at A and A_1, but at all other points its surface lies outside that of the sphere ; in fact, the ratio of the major axis PB to the minor axis PA is the same as the ratio of the constant refractive index for the ordinary ray (1·658) and the minimum refractive index for the extraordinary ray (1·486). In Fig. 194 this ratio has been somewhat exaggerated for the purpose of constructing a clear diagram.

4. POLARIZATION BY DOUBLE REFRACTION

The importance of the study of polarization was considerably increased when **Huygens** made the discovery that the two rays, the extraordinary as well as the ordinary, were completely polarized in planes mutually perpendicular to each other.

In passing, it may be mentioned that the above fact forms the basis upon which many pieces of apparatus used in studying polarized light have been designed. It is only necessary to cause one of the rays to be deflected, or removed in some other way, in order for the remaining part of the light which emerges from the crystal to be completely polarized. Several minerals, like tourmaline, possess the property of being able, for transmission, either to diminish the intensity of one of the rays very considerably, or else absorb it entirely. A plate of such a substance constitutes a very simple polarizer ; a second similar plate can be used for the analyser.

But let us return to the more complete explanation of this phenomenon. Huygens was absolutely perplexed by these new facts. He assumed, as we have already mentioned, that light consisted of longitudinal waves, which always present the same aspect in any direction perpendicular to the beam. Yet here he was confronted with light in which this symmetry was lacking. A solution to the problem again became an urgent necessity when **Malus** (1808) observed the phenomenon of polarization in reflected light. Again is it to **Fresnel** that we owe the solution. He first showed that the extraordinary and ordinary rays could not interfere with one another. But when they were both allowed to pass through another polarizer whose plane was inclined at an angle of 45° to the planes of polarization of the individual rays, they acquired the property of being able to produce interference. From the very beginning it was quite clear to Fresnel that this phenomenon could not be reconciled with the prevailing hypothesis that light waves were longitudinal. But it was not until 1821 that he openly expressed this view. The explanation of double refraction which he then gave was as follows. Light is propagated through a doubly refracting crystal with velocities which may be indicated by a sphere and an ellipsoid of revolution as in Fig. 194. Spherical propagation only takes place for a light vibration executed perpendicularly to the principal or optic axis, and its velocity of propagation is the same in all directions. Now in a crystal of calcite it is only those directions which are perpendicular to the optic axis that are equivalent. When therefore a ray travelling along some direction other than that of the optic axis is incident upon the crystal, there is one particular direction of vibration of this ray which is distinguished from all others, in that it lies in the plane perpendicular to the optic axis. This is the direction of the line of intersection of the plane of symmetry in the crystal with a plane perpendicular to the ray. This latter plane may be called the plane of vibration since all vibrations of this ray take place in this plane. Now to the extent that some of the vibrations in the incident light take place in the direction of the line of intersection of the two planes they also take place in the plane of symmetry perpendicular to the optic axis of the crystal; these vibrations will be propagated in all directions with the same velocity—they constitute the ordinary ray. The extraordinary ray vibrates perpendicularly to this direction, in the plane of vibration, but in general not in the plane of symmetry. Hence all vibrations incident upon the crystal can be resolved into these two perpendicular components. Since the two component rays travel with different velocities, they will soon become separated and diverge.

The above theory was able to account for the many wonderful colour effects which **Arago** had discovered in 1811. Historically they are very important, since they occasioned Arago and Fresnel to investigate interference phenomena. The results of this investigation at once confirmed Huygens' principle and the wave theory of light. By assuming transverse waves instead of longitudinal ones, as Huygens had imagined them to be, the whole subject was brought to a close for the time being.

In order to produce these colours it is only necessary to place a thin piece of doubly refracting material between a polarizer and an analyser. If the polarizer and analyser are rotated so that no light emerges from the latter, then when a piece of gypsum or mica is placed between them, many beautiful and bright colours are to be seen.

These phenomena remind us very forcibly of similar phenomena produced by thin films, and due to interference. Now after Fresnel and Arago had discovered the conditions under which polarized light can interfere, the explanation of these colours was at once apparent to them. The light coming from the polarizer is polarized completely; it then encounters a doubly refracting material (mica). There, in general, the light vibration will be resolved into two components of which one passes through the mica as the ordinary ray, whilst the other constitutes the extraordinary ray. Both rays then pass to the analyser, which can only allow light vibrating in one direction to pass through it. Since, however, neither of the component vibrations of the light approaching the analyser will have its direction parallel to the principal plane of the analyser, only that part of the components will be allowed to pass which can be resolved along the latter direction. These two resulting portions will now produce interference when they are brought into the plane of vibration of the analyser. But the velocities of the ordinary and extraordinary rays are different, which means that they each execute a different number of vibrations in traversing the thickness of the mica sheet. When they emerge a phase difference will exist between the two rays and this difference will depend on the wave-length of the incident light. If for any particular colour this difference amounts to an integral number of half wave-lengths, darkness will be produced. For other colours it will equal an integral number of wave-lengths; when this happens these colours will be strengthened. But when some of the prismatic colours are removed from daylight the emergent beam will be coloured; these colours will change with the thickness of the mica sheet. They will also change when the mica or polarizer is rotated.

CHAPTER V

LIGHT AND COLOUR

1. Natural Light

THE various properties of diffracted and polarized light, although they are only observed under special circumstances, are nevertheless sufficient to enable us to make many deductions concerning natural light. Those deductions which we are able to make from all the facts mentioned previously will now be briefly discussed. By the term natural light we refer for the time being to all the light waves which are emitted from a white hot body or the sun. The term cannot be defined absolutely. The light from the sun is, as we have seen, different from that of an arc lamp, and even scattered daylight proves to be partially polarized when carefully examined. For the time being we shall neglect these minor details and consider the more important properties of natural light. Since we have seen that all light consists of transverse waves, different rays of light, irrespective of a variation in intensity, can only differ in two respects —their state of polarization and their wave-length.

The State of Polarization. We know what polarized light is, and that ordinary light is generally unpolarized. This means that ordinary light presents the same aspects on all sides, so that however a polarizer may be orientated there is always some vibration present in the light which can be transmitted. On the other hand, we do not imagine that each point in a light wave is vibrating in all directions at the same time. Moreover, we know that light can only interfere when the vibrations form a coherent train of waves, to which statement we have also to add that these vibrations must take place in one plane since vibrations in planes at right angles cannot interfere. Now two coherent rays can interfere even when they have traversed paths different by as many as 2 million wave-lengths. During this time the vibrations must have continued to be executed in one plane. If we assume light having a frequency equal to 600×10^{12} vibrations per second (blue light), then 2×10^6 vibrations will be made in $1/(300 \times 10^6)$ sec. For this fraction of a second the light continues to vibrate in one plane. Do facts contradict these conclusions? Most certainly not. For from the first experiment we can only conclude that wherever we have placed a polarizer, the incident light, *as far as we have been able to observe*, has never been lacking in vibrations which the polarizer could transmit. Now in order that we may make visual observations, it must exist for a much longer period than $\frac{1}{100}$ sec. If therefore the polarizer transmits light continuously, this fact only means that there are no interruptions in the process of transmission approaching 0·01 second in duration.

As far as compatibility with interference observations is concerned, the direction of vibration may change 300×10^6 times a second. The direction of vibration might traverse the whole extent of the compass, and if this were to take place regularly degree by degree, so that after

180 such steps the earlier direction were again reached, this whole sequence of directions of vibrations could be executed $1{\cdot}6 \times 10^6$ times in each second. To this we must add that, in the observation of polarization, the vibrations are always resolved into two component vibrations, and we could detect *no* vibration in a given direction, *only if*, during the whole time of observation, vibrations strictly perpendicular to this direction were the only ones available. Thus, if we are able to establish equality of vibrations in all directions, this means that, if the *whole* of the vibrations are resolved into two component vibrations in mutually perpendicular directions, then as far as we are able to detect, the summations of these are equal to each other.

We have therefore to adopt the idea which thought and experiment indicate, viz. that the vibrations occur in all planes orientated at random, if the whole is considered macroscopically, but that each individual vibration occurs in one definite plane, *i.e.* it is polarized. If, like Maxwell's demon, we were able to observe the individual light vibrations, we should find in a period of about one trillionth (10^{-18}) of a second a light ray behaves very similar to a vibrating tuning fork, which as it hangs freely from a cord is pushed, at intervals of about 100 seconds, first in one direction, and then in another.

The Period of Vibration. It is in consequence of the varying periods of vibration for different light waves that their wave-lengths differ. For if ν is the number of vibrations per second and c the velocity of light in empty space, then the wave-length varies according to the equation $c = \nu\lambda$. We have now to consider the simultaneous and unrestricted propagation of so many different waves through space. A similar problem arose when we were discussing sound waves ; the solution was then made possible by means of a simple experiment. The resultant effect of several notes sounding simultaneously may, for example, be engraved upon a gramophone record, the resultant wave arising from the superimposing of all the component waves. With light waves the variety is infinitely more complicated We have only to remind ourselves that, for example, a white hot body emits radiations of all wave-lengths which, in general, may be measured. Moreover, transverse waves may be polarized, so that even for light of the same colour and intensity it is possible for the plane of vibration to be one of an infinite number. But, in spite of this greater multiplicity of possible conditions and the inconceivable speed of light, we have to realize that, even here, all the waves superimpose themselves on one another. The infinite multiplicity of the state of polarization, however, does not increase the difficulty of the problem, but rather tends to facilitate matters, since there is now an infinite number of planes in which the waves may become superimposed. In consequence of this a further assumption is necessary, viz. that when light passes through a grating, when it encounters a lens, or even when it travels through space, where the light of different wave-lengths is either propagated in different directions or with different velocities, the individual component waves are again isolated and propagated separately. This is a necessary consequence of the fact that light can be propagated in an infinite variety of states through space and transparent bodies.

The above ideas are substantiated by Bohr's theory concerning the excitation of light radiation, for the propagation of waves is very closely related to the way in which they are produced. We should never be able to ascertain anything concerning the production of light waves if it were not reflected in their propagation. The emission of light waves, for example, from a *Geissler* tube filled with hydrogen, occurs with that same utmost irregularity which is characteristic of the gaseous state. For it is owing to collisions between the gas molecules, electrons, and ions, that the extra-nuclear electrons belonging to an atom are caused to jump from one orbit to another, and thereby, in an intimately related manner, to stimulate the emission of light waves. It therefore follows that all these waves, sent out by the various atoms, must be superimposed upon one another. When different atoms send out identical waves, these will unite together and thereby increase the intensity of the radiation. It is this whole array of superimposed vibrations which is received by our optical instruments, and then by our eye, where they enable us to form a picture of the external world around us.

2. COLOUR

The colour of light is determined by its wave-length. But when we speak of colour we always include the perception of light by our eyes. The connexion is a very intimate one, but we are not able simply to identify the colour which the eye perceives with colour as defined in physics. When a physicist speaks of colour he refers to those colours which are produced when sunlight passes through a prism, or is diffracted by a grating. These prismatic colours are, in all respects, completely defined by the number of vibrations made per second by the light, or by wave-length in vacuo. As suspected by **Euler** (1707–1783), the various colours can be distinguished from one another like musical notes of different pitch. When a colour is specified by a definite frequency, it has been defined in a manner capable of only one meaning. We are therefore able to distinguish colours by their frequencies much more exactly than with our eyes.

The physical and physiological colours, however, do not contradict one another, for wherever we have ascribed to light a definite period, a normal eye perceives the colour in question. But even normal eyes are not able to distinguish between colours, which, when examined spectroscopically, are found to differ slightly in their periods. Moreover, our eyes can distinguish two colours—black and white—which are unknown in physics. There is no possible doubt whatsoever about these colours, for we see them just as clearly and distinctly, and in the same manner as, for example, we see red and green. Again, we very often say we see one colour, when physics can prove quite definitely, that our eye is receiving waves of very different wave-lengths. Rays which we could perceive, and always do perceive, when they are examined individually, together produce a perception in our eyes which may be very different from that of either single colour. If, for example, yellow and blue colours are received by the same portion of our retina, the sensation of white is produced. In addition, we have the fact, that very often we are not able to notice all the differences which a physical investigation would reveal to us. For example, when red and green rays fall upon the retina of our eye simultaneously the sensation of white is produced. This same sensation of white is also produced when yellow and blue rays are seen at the same time. A prism would reveal the two colours at once, a thing which we cannot do, for however often and closely we may examine the colour, it will always appear to be white.*

---

* These remarks, like others in the present section, are only true if the two colours are *not* pure. [*Trans.*]

From the above-mentioned remarks, we see quite distinctly that we are here dealing with two fields, which, although they are connected in many ways, and depend upon one another, may nevertheless be confused unless they are considered very carefully.

Objective and Subjective Factors in the Perception of Colour. From these and other facts we derive the conclusion that our eyes themselves are an important factor in the process of seeing. Our task is therefore to discover by close examination what it is that is present in light and in bodies, and what part is played by the eye, in connexion with the perception of colour. The only certain way in which we can achieve this aim is to use such means that our eyes do not play the decisive rôle in the investigation. We are able to do this without much difficulty. We have only to examine, with the aid of any spectroscopic apparatus, the light which falls upon a body of any one colour, and the light which it emits. The results of such an experiment are as follows.

If a piece of red paper, such as is used by photographers in their dark rooms, is held in a solar spectrum, it only appears to be red where the red rays fall upon it—everywhere else it is black. This means that a body only appears to be red (green, etc.) when it emits rays of this particular colour. So long as the body is not self-luminous, it must, if it is to exhibit colour, be irradiated with light of such a colour as it can reflect. When the question is raised concerning the source of energy which enables a body to appear coloured, we have to reply that it comes from the sun or some other luminous source. From the light which bodies receive, they select only a portion, which is reflected to us so that the body appears coloured. Whilst a body continues to emit light of one definite colour our eyes only perceive this same one colour.

When a piece of white paper is held in such a position that a spectrum falls upon it, the various prismatic colours are seen in their proper order and sequence. A body will therefore appear white whenever it returns to us all the incident rays which fall upon it ; but these rays must not come from a smooth reflecting surface, but proceed from points in the surface of the body, from whence they are scattered in all directions. If the body returns none of the incident radiation it is said to be black.

The remaining rays which fall, for example, upon a red body, become absorbed in the interior of the body. Bodies reflect a portion of the incident radiation from their surfaces, but this light consists of all wavelengths, so that, in general, it is without colour. This portion increases as the degree of polish on the body is improved. But the more perfect the polish becomes, the less pronounced are the colour characteristics of the body impressed on the reflected light. Light, therefore, which is reflected from a perfectly smooth object, will only impart to us information concerning the source in which it had its origin, and not about the object it has encountered. In general, however, some of the light passes below the surface, where all bodies absorb light of some definite wave-length, but transmit a portion of the remaining light, and the rest, of course, is reflected. The transmitted light penetrates into the lower layers of the body, and, at each new molecule which it encounters, part will be reflected and part transmitted. If the light contained waves, which had not been absorbed in the upper regions, but which are capable of being absorbed by the body in question, then the percentage of such light will diminish as the penetration increases. We have to remind ourselves that bodies do not fill space completely, so that a portion of the incident light will be absorbed only in the lower layers of the object. The reflected light will therefore have its origin in the lower layers of the body, and it is this light which determines the colour of the object. Even with transparent but coloured substances, a portion of the light which encounters them is absorbed, the rest being partly reflected, and partly transmitted. It is owing to this that a piece of red glass appears to be red, whether examined by transmitted, or by reflected, light.

The various colours which different bodies are able to absorb or reflect are a characteristic feature of these substances, and they are responsible for their body colours. One can easily realize the vast array of such colours which are possible, for it is only necessary for the body to absorb **one** definite colour, in order for its body colour to be different from that of others. The impression which all the reflected rays make upon our retina is the colour we assign to the body. We have now to discover whether or not we are able to perceive the combined effect of several light waves.

As already stated, our eyes are not capable of resolving light into its component colours—the impression which we receive is always a resultant effect. In the first place, there is a whole series of colours, which can be arranged in pairs, so that white light is produced whenever they are superimposed. We refer to the so-called **complementary** colours, red and green, yellow and blue, yellowish-green and violet. On the other hand, red and violet produce a colour, which we call purple. **Newton** attached variously coloured discs to a plate. When this was rotated rapidly the individual colours could no longer be distinguished ; the colour observed is a combined effect due to all the colours of the various discs. Later this apparatus has been improved in many details, but we cannot discuss them here, for they belong to physiology, rather than to physics.

In contrast to these summation colours we have to distinguish the so-called residual colours (*Restfarben*), which we shall proceed to discuss briefly. They are always produced when two body colours, *e.g.* oil paints, are mixed with one another. Let us assume that yellow and blue pigments have been intimately mixed together. The artist knows that, under these circumstances, green will be produced, although by superimposing a yellow and blue impression they appear white to our eyes. When these two pigments are mixed, and thoroughly stirred, the smallest yellow and blue particles still actually remain very close to one another, without amalgamating. If white light falls on such a mixture the yellow pigment will absorb blue rays in preference to all others, the red, yellow, and green rays being partially absorbed to varying extents. The light which is reflected from the yellow particles will encounter some of the blue ones. These absorb the red and yellow rays, so that the only light, which is finally reflected from the mixture, is that which is not absorbed by either of the pigments, viz. green. The green colour which is characteristic of such a mixture is never so clear, or, as one says, so rich, as that of a single homogeneous green dye. This is in complete harmony with the explanation which has been given, since many of the light vibrations are reflected after they have encountered only one of the two component pigments.

CHAPTER VI

THE ÆTHER

IN the whole of physics there is no idea which, for several centuries, has raised such difficulties, and which, even to-day, is so much discussed, as that of the æther. It is not that different schools of physicists have proposed various hypotheses, which they refuse to abandon, but, as we have already remarked, it is rather the facts themselves, which appear to contradict one another.

There are some facts which seem definitely to require an æther with quite definite properties, whereas other facts cannot be reconciled with such properties at all. It has often happened, that, on the basis of some facts, a hypothesis has been postulated, and then, as is usual, conclusions have been deduced from it with the idea of testing them experimentally. When the new experiments have been performed the results have frequently been completely out of harmony with the hypotheses. Even to-day the question of the æther has not been solved, and to judge from the difficulties which have hitherto been raised, many years must necessarily elapse before the problem is solved. In dealing with the æther, then, our only course is to set forth the various arguments and counter-arguments, and hope that coming generations may be able to discover the solution to the problem.

1. The Existence of the Æther

The idea that there was a material something which filled all space, and which carried messages from the stars to us, was familiar to **Aristotle**. He maintained very definitely, and in known contrast to the other philosophers of his time, that between the luminous object and the receiving eye there must be a fine substance to carry the light from the object to us. If the intermediate space were a complete void, objects could not be seen.

But we cannot say that Aristotle's hypothesis was founded on scientific lines, for the ideas which were then prevalent concerning the tasks which the æther had to fulfil, and especially about the nature of light, which the æther served to transmit, were very imperfect. The idea of a material something having definite properties, so that it could fulfil certain definite tasks, originates from the time of **Newton**.

It is quite clear from his writings that Newton was an avowed opponent of the idea of action at a distance. He assumed that the attraction between two masses only took place owing to the presence of a very rarefied substance, which was between them. A paragraph, at the conclusion of his treatise on mechanics, begins with the following words: 'Even now, something could be said concerning an extraordinarily fine æther (spiritus), which lies between all bodies, and which even penetrates into their interiors. **It is owing to this æther that material particles attract one another,** and that electric charges attract or repel one another. . . . But we lack a sufficient number of facts, from which the laws governing the mode of its action can be determined exactly and proved.'

The following are, in short, the present-day reasons which cause us to assume the existence of such a material substance, generally termed the æther.

1. **Gravity.** As already mentioned, we have no positive indications of the manner in which gravity acts, how it is propagated, or concerning its velocity. The reason why we assume that gravity requires a medium is because we are convinced that action at a distance is impossible. The validity of this hypothesis has been confirmed again and again in the course of further investigations, for it has been found essential to postulate, for all other phenomena, which are propagated through space, a medium which participates in the action concerned. It is impossible for us to conceive how, for example, the earth could be attracted by the sun, if there is no intervening medium present. The action of the sun could not be such that it suddenly emerged from nothing, seized our earth, and caused it to move in its orbit around the sun.

2. **The Propagation of Light.** It could quite easily be imagined that, in Newton's emission theory, the æther was a superfluous conception. Actually, however, from the concluding section in his treatise on mechanics, to which reference has already been made, Newton required the presence of an æther. This necessity arose for the purpose, amongst others, of emitting light, of reflecting, refracting, and diffracting light, and of heating bodies. Perhaps he had assumed that the light corpuscles and the æther were identical, and that such particles, present in bodies, exert an influence upon the light corpuscles which encounter them. When light had been shown to consist of waves, the presence of a material substance, which carried the light from the sun to us, or from star to star, became inevitable and unavoidable. It is impossible for a wave motion to arise unless there is a medium through which the motion can travel. This medium is, temporarily, the bearer of an enormous amount of energy passing from the sun to us, which is afterwards given up to the earth, where all terrestrial life and actions are sustained by it. All wave theories postulate the existence of particles which are constrained to some zero position by forces ; when the particles are displaced from such positions it is these same forces which cause them to return. It is also assumed that these particles possess inertia, so that on their return towards the zero position, they shoot beyond, thereby giving rise to an oscillatory motion. It does not matter, for this purpose, whether the forces are elastic or electric in origin, or whether the inertia is mechanical or electromagnetic.

3. **Even electrical attraction and repulsion** due to charges is propagated across the intervening medium. As we have already seen, it was **Faraday**, in particular, who showed that some medium must have been participating in these effects. For no other action at a distance have we such definite and sure indications that such a medium is present for the action to take place, as for those which Faraday provided for us concerning electrical phenomena [*cf.* p. 242]. It is true that the attraction between two charged bodies is propagated through the intervening space, just as well as the electrical forces, which are the primary cause of electrostatic induction [*cf.* p. 238] ; but when the hypothesis is restricted so that only essential things shall be retained, it does not appear necessary for us to assume that the two actions will be separately conveyed through space. It suffices to assume an electrical displacement which is directly responsible for the attraction and repulsion in bodies. Electric and magnetic forces are also propagated through completely air-

free space, and their effects are so little altered as a complete vacuum is approached, that here also we can only ascribe a subordinate rôle to the bodies which may be situated in the field. We need an æther which is also capable of transmitting electric and magnetic actions.

For a long time physicists were accustomed to assume the existence of several ‘imponderable fluids’. Positive and negative electricity were assumed to be due to the presence of two such electric fluids, and even heat was supposed to be a substance known as ‘caloric’. These fluids were considered to be the media which were responsible for the transmission of the various effects through space. But since it is now known that heat is merely a form of energy, and that radiant energy is not different from light waves, the caloric theory has been entirely dropped. Also the electricity in charged bodies, which consists of an accumulation or a deficit in electrons, is something essentially different from the æther. By this we do not necessarily mean that electrons play no part in the æther,—on this point we know nothing.

Finally, it has been shown that light and electric waves are fundamentally the same, and that magnetic forces always arise whenever electricity is in motion. Hence, the requirements (2) and (3) become welded into one for several reasons, all of which, however, only require one and the same medium. There are therefore only two essentially different reasons why an æther should be postulated. It is needed in order to account for the gravitational attraction between matter, and for the propagation of electromagnetic waves. This medium fills all space, or at least all space as far as the remotest star from which we receive light radiations.

2. THE PROPERTIES OF THE ÆTHER

In order for the æther to fulfil the tasks which have been imposed upon it, it must possess certain characteristic properties. Such a statement appears to be so self-apparent that it is difficult to realize the difficulties to which it has led.

First and foremost the æther is different from all other substances which we perceive with our senses, or which we can place in various receptacles. It does not belong to the periodic system of the elements. We have to look upon it, however, as a material substance, for it is very intimately associated with matter. The motions of the atoms stimulate vibrations in the æther so that energy can be taken from a body and transferred to another. The enormous amount of radiant energy which is responsible for all life and industry on this planet is, as it travels to us from the sun, for eight minutes in the sole possession of the æther. The principle of the conservation of energy must apply to the æther, for otherwise it would not be valid in connexion with the exchange of radiant energy between two ordinary bodies.—Does the æther possess inertia, and does it succumb to the action of gravitational forces? This is constantly denied, and actually we do not know of any process from which we might establish with certainty, or even with any degree of probability, the fact that the æther is so endowed. At the same time we must remember that our inability to detect a thing does not prove its non-existence. As long as we know nothing as to how the æther takes up the solar energy, it is too early to affirm anything either one way or the other. It is true that we know that energy comes to us in the form of electric waves, but it is quite unknown to us of what the electric waves in the æther consist.

We only know that the waves in the æther are transverse. Hence, whatever medium there may be, it must be capable of taking up and transmitting transverse vibrations. We have often made the statement that it is only solid bodies which have this property ; the æther which we postulate must therefore possess the properties of a solid body. That appears to be inevitable. But if interstellar space is filled with a solid substance it appears impossible for the celestial bodies to continue their courses for thousands of years through this space without experiencing a considerable decrease in velocity. The requirements of the propagation of light and of the movements of the celestial bodies appear to stand in direct contradiction to each other. This difficult position was only slightly alleviated when **Stokes** pointed out that the velocity of light is enormous in comparison with the velocities of these bodies, and that a medium may appear to be solid with respect to a body which is moving rapidly through it, whilst to a more slowly moving body the medium may behave as a fluid. He reminded us that such substances as pitch can be crushed by a blow, just like any brittle substance, and yet if they are observed for some time, they will be observed to be flowing slightly outwards. The resistance which a ship's propeller experiences when it rotates rapidly is so great that even a powerful man-of-war can be driven quickly across the ocean. But water only offers a very small resistance to objects which move slowly through it. Similarly the resistance of the air, which is usually negligibly small, becomes very great when the propeller of an aeroplane is set rotating ; it is sufficient to cause the aeroplane to move forward with considerable velocity. In comparison with the velocity of light, however, the velocities of the heavenly bodies are very small indeed.

At the time when only longitudinal waves had been recognized, it was thought that light waves must necessarily be similarly constituted ; even the density and elasticity of the æther were calculated. The velocity of propagation is expressed by equation (46), viz. $c = \sqrt{e/d}$. Moreover, the energy which the æther receives from the sun is also known. If some probable value is chosen for the velocity of the æther particles, then from these two equations the density and elasticity can be derived. The order of magnitude of the density is 10^{-20}. It is somewhat ludicrous to discuss this question further, since in view of the fact that light waves are electromagnetic in origin, the reasons which led such a calculation to be made are really very feeble.

It might be thought, perhaps, that this knowledge would remove all difficulties. But this is not so. Even the electric charges which we now assume to be present in the æther, must always be constrained by forces tending to hold them in some zero position, *i.e.* they must be linked to neighbouring charges.

When physicists found themselves confronted with these difficulties, it is not surprising that they should always have been on the *qui vive* for new facts which would help to reveal the properties of the æther. But such experiments proved themselves to be enormously difficult. They met with very little success. Let us consider the matter.

We determine, for example, the weight of a body. We do this by observing the difference in the extension of a spiral spring when it is loaded with the object in question, and when the body is removed. But we cannot do this with the æther, which is ubiquitous and cannot be removed from any one position. Or we may examine the shape, surface, or colour of a body. In order to do this we observe the difference in the light which comes from a certain spot, first when the body is there, and secondly when it has been removed. Such experiments, however, cannot be made with reference to the æther, for the æther cannot be removed. But it must not be concluded that, because we cannot detect the effects due to the æther, such effects do not exist, and that the æther does not participate in any process. Rather we are convinced that it intervenes in the propagation of gravitational forces, and that it brings to us that solar energy which is so necessary for all forms of life. But we are not able to show either qualitatively or quantitatively that this intervening action does occur. Usually we show whether any particular circumstance plays a part in a phenomenon by observing the effect of increasing or decreasing its influence, or even of cutting it out altogether. But all this is not possible to us when we are dealing with the æther.

3. The Motion of the Æther

Fizeau's Experiment. If the æther is the carrier of light waves, then one would expect that, if the æther moves, a corresponding change in the velocity of light would be observed. At any rate the French physicist **Fizeau** decided to investigate the matter and find out whether or not the æther was affected when a body moved through it. If this does happen, then the velocity of light will be different according as the light is transmitted in a direction opposite to that of the moving body, or vice versa. Since it is to be expected that the changes in velocity will be very small, instruments of an exceptionally sensitive nature will have to be devised. Interference methods are such that these small changes in velocity may be detected, for a shift equal to the width of one fringe corresponds to a path difference of half a wave-length. This corresponds to a time difference of less than 10^{-15} sec.

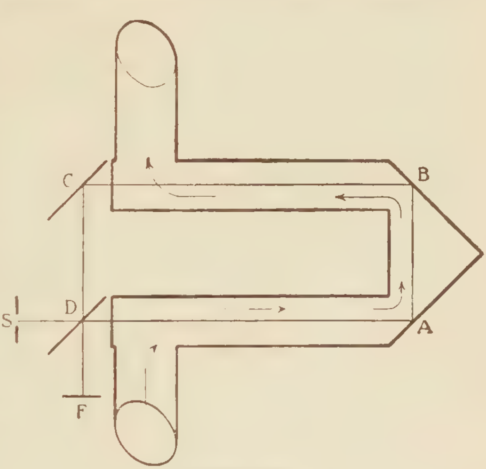


Fig. 195.

Fizeau's apparatus is shown in Fig. 195. Rays from an illuminated slit S were incident upon a half-silvered mirror D. A portion of the light passed along the course $DABCD$ and eventually arrived at F: a second portion traversed the path $DCBAD$ before it also arrived at F. Interference bands were produced at F. The path which these light waves traversed was through two tubes which could be filled with water or any other liquid. By means of a powerful pump this water was driven with considerable speed in the direction of the arrows. It is easily seen that the ray $DABCD$ travels with the water, whilst the ray $DCBAD$ travels against it. If now the water in virtue of its motion carries the æther, and therefore the light waves, along with it, then the waves will be lengthened when the light passes in the direction of the running water: the position is similar to that of an athlete who is able to jump further with the wind, than against it. The light will therefore pass more rapidly through the moving water. This must necessarily happen since all wave-lengths have been increased so that the number of waves actually present along the path will be less than before, *i.e.* the time of transit is shortened. On the other hand the light waves which move against the water will be shortened so that a greater time will be occupied in traversing the distance along the tubes. If, for example, the time difference when the water is moving is increased by only one half-period, instead of a bright fringe a dark one will be seen at F. All this is an immediate consequence of the assumption that the æther is carried forward with the water. If the æther is not disturbed by the running water, then no change will be observed in the interference bands at F. In the above instance it is even possible to calculate the shift which will be obtained when the water is moving with a definite velocity.

The result of this experiment was very surprising; the interference fringes were displaced, but not as much as had been calculated. The experiment was repeated several times, the same result always being obtained. It is therefore necessary to conclude that the æther has been carried forward with the water, but not with its entire velocity. The degree of transport proved rather to be dependent on the index of refraction of the liquid, and if the calculated change of velocity be 1, it was found experimentally to be $\left(1 - \frac{1}{\mu^2}\right)$, for a fluid of refractive index μ. It was thus reduced, but nevertheless it could be easily observed for water, where $\mu = \frac{4}{3}$. With air, of refractive index practically unity, the æther is not dragged along to any appreciable extent, and this result was confirmed experimentally. Hence the final result is that with air the æther is *not* carried along.

Michelson's Experiment. Michelson, however, made further deductions from these experiments. According to him we are moving with the earth against a stationary æther. Hence one might attempt to ascertain whether or not the moving earth experiences a resistance or some other effect due to the æther, and thereby show the relative motion of the earth with respect to the æther. In short, we should have to experience an 'æther-wind'. This term is, of course, only a pictorial representation of the state of affairs which may exist. In the previous instance it would indicate a difference in the velocity of light, according as it was moving with or against the æther. Hence if a suitable experimental arrangement could be devised, and the experiment then carefully performed, results might be obtained. Michelson devised the experiment which has been very frequently quoted during recent years. He endeavoured to prove that the earth did move with respect to the æther, by measuring the velocity of light. His reasoning was as follows.

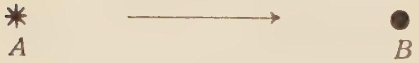


Fig. 196.

Perhaps it will be more expedient if we develop his line of thought by means of an example. Let A, Fig. 196, be a gun station from which a shot is fired at a definite instant. Let B be the position of an observer who measures the interval of time between the firing of the gun and the arrival of the sound at B. If the distance AB has been measured he can calculate what this time should be; for our present purpose we can neglect the time taken by the light to travel from A to B. Will the results of the observer at B agree with those that he has calculated? Only when the air between A and B is at rest. If the wind is blowing towards B the sound will arrive sooner than the observer had expected, for in addition to the condensations and rarefactions in the air, which are responsible for the propagation of the sound, the air itself is moving forward. If the observer was not cognizant of the fact that there was a wind, he would find that the velocity obtained for the sound would be too great. If c is the velocity of sound in still air, whilst v is the velocity of the wind in the direction AB, he would assign a value $c + v$ to the velocity of sound according to his determinations. We may call this the apparent velocity. We say that it is the apparent velocity, since we presume that we already know the true velocity, and that the motion of the air, *i.e.* the 'wind', can be perceived in some other way. But if the observer at B were not acquainted with these facts, he would simply assert that the velocity was $c + v$ without making any other comment.

But if the experiment were repeated with the wind blowing in the opposite direction, the apparent velocity of the sound would be $c - v$. If therefore the apparent velocity in the first direction is a, *i.e.* $a = c + v$, whilst in the second it is b, where $b = c - v$, then the relative velocity of the air with respect to the earth can be deduced; it is in fact given by

$$a - b = 2v.$$

Similar considerations can be applied to the propagation of light. If the æther is carried along with the earth, then there will be no æther wind between A and B, *i.e.* the light will travel from A to B, or B to A, with the same velocity as in a stationary æther. If the velocity of light between these two points were measured, let us say with an apparatus similar to Foucault's, then the true value of the velocity would be obtained. But if the æther is not carried forward by the earth, then conditions between the two stations are just as if an æther wind were blowing. If, for example, the earth, with the light source at A and the receiving apparatus at B, moved forward in the direction AB without carrying the æther along with it, there would be an æther wind blowing from B to A. Any messenger travelling from A to B would have to overcome the effect of this wind, and therefore he would require a longer time in which to make his journey. The return journey would of course be made more quickly. From these two times the velocity of the earth with respect to the æther could be ascertained.

The above paragraph is an epitome of the thoughts which led **Michelson** to contemplate an experiment on these lines. The real difficulty arose when an attempt was actually made to perform the experiment. Now the greatest service rendered by Michelson, in connexion with this problem, was not that he had formed the above ideas, but that he devised an experimental arrangement by means of which they could be tested. The method was such that very small differences in the velocity of light could be measured very accurately. For this purpose Michelson made use of the phenomenon of interference.

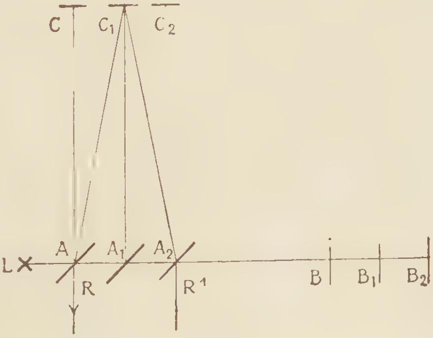


Fig. 197.

Light from a source L, Fig. 197, and which only contained waves of one definite wave-length λ, were caused to fall on a half-silvered mirror A which was inclined at an angle of 45° to the light. One portion of the light passed through the plate A and encountered a mirror B, which reflected it along its original path, so that it was again incident upon A, and some of this light was reflected towards R.

The other portion of the light reflected from A travelled to C, from whence it was reflected so that it again fell upon A, and then passed to R. Since the two paths AB and AC had been made equal, it follows that each will contain the same number of wave-lengths, *i.e.* the two portions of the incident light which arrive at R will produce interference bands.

If one of the two light rays experiences an æther wind, which, let us say, retards it, it will arrive somewhat late at A. In consequence of this the two wave trains which arrive at R will no longer be in phase. If the delay amounts to one half-period of the light vibration, then a bright band at R will be replaced by a dark one ; in fact each bright fringe will be shifted into a position originally occupied by a neighbouring dark one.

Such a displacement will also occur if the whole apparatus is rotated. Let us assume that the earth is moving around the sun with a velocity of 30 km. per sec. in the direction AB. The ray AB will experience an æther wind, whilst the other AC which is perpendicular to AB will not be detained by this wind. But when the whole stone table, on which the apparatus was constructed, is rotated through 90°, the ray AC will experience the wind. The first ray will no longer be delayed, but the second one will. Hence by rotating the apparatus twice the original shift could be produced.

The delicacy of the experimental arrangement used by Michelson was such that he could refer the change in velocity to the change in phase of the light waves arriving at R, which he could calculate from the accurately known shift of the interference fringes which might take place. One may perhaps raise the objection that when the light waves travel in the reverse direction (*i.e.* after being reflected from the mirrors B or C), the difference which they might have experienced will be nullified. To this we have to reply that any effect which is produced before the reflexion takes place will be partly eclipsed after the reflexion, but the resultant effect is still large enough to be observed. The following argument will show that this is true.

According to the result of Fizeau's experiment Michelson regarded the æther as being at rest, so that the earth together with the experimental arrangement which he had built would move through this stationary æther. Let the light be emitted from A at the time $t = 0$, with a velocity c, and suppose it arrives at B, where $AB = l$, at the time t. Then $l = ct$. Now let us suppose that the earth moves with velocity v through this stationary sea of æther. The light will no longer encounter the mirror at B because, during this time, the mirror will have moved forward with the earth. Consequently a longer time t' will have elapsed before the light reaches this mirror. During this time the mirror will have been displaced by an amount BB', which is expressed by $BB' = vt'$. The whole distance from A to B' is therefore $l + vt' = l'$ (say) ; and the time, t', which the light needs to travel across this distance, is $t' = l'/c = (l + vt')/c$.

Hence $t' = \frac{l}{(c - v)}$. The time is therefore greater than when the earth is at rest.

After reflexion at B the light will eventually fall upon A. The time will be shortened. Let this new time be t'', so that A will have advanced a distance vt''. The distance across which the light has travelled will therefore be diminished by vt'': its total path is represented by B_1A_2. Let this be $l'' = l - vt''$, so that the time taken to travel across this distance is

$$t'' = \frac{l''}{c} = \frac{l - vt''}{c}, \text{ or } t'' = \frac{l}{c+v}.$$

The total time taken for the journey from A to the mirror and back again to A is therefore

$$T = t' + t'' = \frac{l}{c-v} + \frac{l}{c+v} = \frac{2lc}{c^2 - v^2} = \frac{2l}{c}\frac{1}{(1-v^2/c^2)} \quad . \quad (196)$$

This formula shows that the time of transit is different when the earth is moving, for when the earth is at rest the time is $t = \frac{2l}{c}$.

Now let us follow the other portion of the light which has travelled to C. A glance at the right-angled triangles AC_1A_1 and $A_1C_1A_2$ shows that the paths will again be slightly different. If $AC = AB = l$, then the total path can be expressed in terms of l and v, and a calculation, similar to the one we have just made, will show that the total time for the light to pass across to C_1 and back again to A_2 is

$$T' = \frac{2l}{c} . \frac{1}{\sqrt{(1 - v^2/c^2)}} \quad . \quad . \quad . \quad (197)$$

In consequence of the fact that T and T' are different, a shift of the interference fringes will take place when the whole apparatus is rotated. The total delay of one ray with respect to the other is $T - T'$, so that when the table on which the instrument stands is rotated through 90°, the shift will be proportional to $2(T - T')$. Since $\frac{v}{c}$ is small we may use the more approximate expression

$$2(T - T') = \frac{2l}{c} . \frac{v^2}{c^2},$$

so that the difference in path is

$$\Delta l = 2c(T - T') = 2l . \frac{v^2}{c^2} . \quad . \quad . \quad . \quad (198)$$

Now whenever this difference amounts to one whole wave-length of the light which is being used, the interference bands will be displaced by an amount equal to the distance between two consecutive bands. Hence the total displacement is

$$\frac{\Delta l}{\lambda} = \frac{2l}{\lambda} . \frac{v^2}{c^2} . \quad . \quad . \quad . \quad . \quad (199)$$

when expressed in terms of the distance between two bands.

In Michelson's experiment $l = 11$ *m.* $= 1100$ cm. The light used was that of the red cadmium line, so that $\lambda = 5{\cdot}9 \times 10^{5}$ cm. Since the velocity of the earth around the sun has an average value of 30 km. per sec., we have $v/c = 10^{-4}$.

One would therefore expect that when the whole apparatus was rotated through 90° the shift would amount to

$$\frac{2 \times 1100 \times 10^{-8}}{5{\cdot}9 \times 10^{-5}} = 0{\cdot}37$$

of the distance between two bright fringes.

The above is what would be expected according to Fizeau's experiment, which was regarded as proving that the earth did not carry the æther along with it in its journey through space. As is well known, Michelson's experiment always led to a negative result; that is to say, however often the experiment was repeated, and however carefully the fringes were observed (one-hundredth part of the expected shift could easily have been detected), not the least influence of the earth's motion with respect to the æther was ever found.

Fizeau and Michelson had therefore each, by means of experiment, proposed a question to Nature. The answer which Fizeau obtained was that the earth moves through a stationary æther, whilst Michelson received a reply to the effect that the earth does not move with respect to the æther. None is felt, not even where it might be expected if it were present; the æther must therefore be carried along with the moving earth. These two replies simply contradict each other. Many physicists in all countries have spent much time considering this question in the hope of detecting some experimental error or flaw in the method of working out the results: but their efforts were always in vain and the two results remained incompatible with each other.

The Significance of the Results obtained by Fizeau and Michelson. In addition to the difficulty which had been encountered when it was found necessary to assume that the æther was sometimes a rigid body, whilst at others it was very attenuated, another now presented itself. According to Michelson the æther moved with the earth, whereas according to Fizeau it did not move with it. One has to confess that science was placed in a very difficult position. For if it were possible for one fact to contradict others, then all the efforts which have been made to acquire knowledge would have been in vain, and we should no longer be able to deduce the simplest conclusion from any fact, since other facts would show immediately that our conclusion was false. How was this difficulty to be avoided? Its removal was the task which **Einstein** imposed upon himself in 1905. The facts themselves had certainly been established, and we have always to adjust our views and hypotheses to suit the facts, and not vice versa. The facts had to remain, therefore, whilst an endeavour was made to discuss what was wrong with our theories.

It was unfortunate that in explaining these two experiments only the most fundamental assumptions concerning light had been used, so that if anything were amiss it would certainly involve these hypotheses. These, however, could not be altered without causing an upheaval, the effects of which would certainly be very far-reaching. It was scarcely realized that it would adumbrate all knowledge which had hitherto been acquired. Even the most prudent physicists became excited when Einstein undertook to bring this indefensible state of affairs to an end, and refused to be turned aside from his set purpose by any of the menacing consequences which confronted him. Einstein maintained that his new ideas were a necessity not only for the further development of physics, but also if it were to continue as a rational science.

The difficulties which had appeared in the traditional hypotheses of physics have to be recognized as such. They formed the nucleus from which Einstein's theory has grown, **but, even to-day, nobody has succeeded in explaining them.**

It is not a question of asserting a previously conceived opinion nor still less a conception of the universe; it is solely a question of shaping our conception of reality, in so far as it is revealed by established facts, so that it is free from contradiction.

The difficulties had arisen in connexion with certain observations concerning the motion of the earth with respect to the æther, for if this motion did exist, it was expected that it would affect the transmission of light. But only negative results had been obtained. The propagation of light waves was the same whether it occurred in a medium which was moving with respect to the æther, or in one which was at rest. Speaking somewhat more strictly, it may be said that the velocity of light was the same in each instance, so that no answer had been obtained concerning the question of the earth's relative motion with respect to the æther. We may remind ourselves of the remarks which have already been made concerning relative motions in mechanics [*cf.* p. 31]. According to Newton's mechanics, it is not possible to make any conclusions concerning the state of the motion of the earth, a train, or a ship, from the observations of an observer who may be present upon them, and equipped with suitable instruments, at least so long as these motions are not accelerated. This was a consequence of the principle of the independence of motions. The effects of a force on a body are entirely independent of whether or not it is here or there, or what may be its velocity. We now see that the same laws are true for light. Light travels from a point A to a point B, both of which may be on the earth itself, with a velocity which is independent of any motion the earth may possess. Einstein saw that these facts were really embodied in a universal law of Nature, according to which it is impossible to ascertain the absolute motion of the earth from any process in the universe whatsoever. He also assumed a second general law which asserts that the velocity of light is the same whether it is measured by an observer moving in the same system as the light source, or in another whose motion is quite different.

It was found very difficult, however, to reconcile these new ideas with those which had hitherto been developed ; in fact, it was quite impossible. Einstein realized that he was compelled either to forego a solution to this problem, or else to introduce conceptions differing in essential details from those which were then held. These supposedly revolutionary ideas occasioned much excitement ; for several years they attracted the attention of many to whom a study of physics makes no general appeal, but now, fortunately, this excitement has given place to serious consideration of the theory by the expert.

4. EINSTEIN'S THEORY OF RELATIVITY AND THE ÆTHER

The problem of the æther was solved by Einstein in a very simple manner : **he assumed that there was no æther.** He was led to make this assertion because, if the æther did exist, then according to **Fizeau** the earth should be moving through an æther which is at rest, whilst **Michelson** required an æther which was carried forward by the earth. Since it is impossible for both these conditions to be fulfilled, the æther cannot exist. In presenting this new theory, Einstein was really asserting that the results of the two experiments were so contradictory that the difficulties could be removed by no other expedient. It is true that these difficulties had now been dispelled, but others arose to take their places. Even the wave-nature of light was refuted, because waves can only be propagated in an elastic medium. Einstein admits all this quite openly, but he hopes to explain the propagation of light by some sort of emission theory—*i.e.* something is emitted from the luminous body which can travel through space. All the objections which had been raised against Newton's emission theory are now revived. For we have measured the wave-length of light, we have observed phenomena due to interference and diffraction, and determined the plane of polarization. All these suggest the presence of waves and therefore require the existence of an æther. It has to be remarked, however, that in the general theory of relativity, which we shall soon proceed to discuss, Einstein has introduced a kind of æther, for he has ascribed certain properties to space. 'The gravitational field is a condition in space, which simultaneously determines gravitation, inertia, and the metrics of the space.' Even if a ray of light passes through such a gravitational field, its path will be altered and its velocity changed. But Einstein has not again assumed the æther to be responsible for the transmission of light waves. From these two postulates, *i.e.* from the theory of relativity and the constancy of the velocity for light, certain relationships can be deduced, for an event (*i.e.* the coordinates of a point and the time at which a particle is at this point) referred to systems which are at rest or moving. It is found that distances become shorter, and that the duration of events becomes less, if they are estimated from a moving system of coordinates. The ratio is the same in each instance; in fact, if v is the relative velocity of the two systems whilst c is the velocity of light, it is equal to

$$\sqrt{1 - v^2/c^2} = l : l' \text{ (say)} \qquad . \quad . \quad . \quad (201)$$

Since the numerical magnitude of c is very large in comparison with all usual values of v, this ratio will only differ from unity in quite exceptional instances. Hence it was not a matter of discrepancies between experiment and observation, but rather the fundamental influence of the state of motion on dimensions, that appeared to be so strange. But the reader should be careful not to use such expressions as, 'a rod is shorter when moving than when it is at rest'; such statements are mere epitomes rather than exact representations of the truth.

The general interpretation which one gives to the above equation is somewhat as follows. If a moving rod is measured according to the rules which Einstein has laid down, then the result is different from that obtained when the rod is measured according to the rules for stationary objects. But if from the length accorded to the moving rod and its velocity, one calculates what its length would be if it were in a stationary system, the same magnitude is obtained as if the length had been measured directly when the rod was in the stationary system. There is no question of any contradictory facts nor yet of any uncertainty. Similar remarks may be made with reference to observations on time.

We have already mentioned one consequence of this new theory: it is the one which Einstein considers to be the most important, and it concerns a relationship between the inert mass of a body and its energy content [*cf.* p. 299]. According to it, if the energy content of a body is increased, its inertia also becomes greater. If the increase in energy is E, then the increase in inertia is given by

$$m = \frac{E}{c^2} \qquad . \quad . \quad . \quad . \quad . \quad (202)$$

where c is the velocity of light.

If therefore the change in energy is not large, the increase in the inertial mass is only very small. Even this relationship is sometimes expressed as if an increase in energy, which may be due to an increase in velocity, causes the quantity of matter in a body to change. The true meaning has already been given.

It is supposed that the fact that the atomic weight of helium is not exactly four times that of hydrogen is due to this relationship between mass and energy. The atomic weight of hydrogen is 1·008 whilst that of helium is 4·00. Now four hydrogen atoms would have a total weight 4·03, which is a little greater than that of one helium atom. But in order to split a helium atom into four atoms of hydrogen an enormous amount of energy would be required; we have already seen that even after an explosion in a radioactive atom, the helium atom emerges unscathed, a fact which shows that the helium atom is more rigid than the remainder of the radioactive atom. By the aid of equation (202) it is even possible to calculate how much energy must be given to a helium atom before a breakdown into four hydrogen atoms could be effected. It is

$$E = 0{\cdot}03c^2 = 27 \times 10^{18} \text{ ergs},$$

an amount which is about one million times that liberated in the most violent chemical reaction known to us. Even an alpha particle would not be able to effect a rupture. In a similar manner it is hoped that it will be possible to explain on the theory of relativity the other small deviations from integral values of the atomic weight, which are still to be found in various positions of the periodic system.

The General Theory of Relativity. If we now ask ourselves whether or not we are quite sure that we cannot detect a motion unless there is some body to which the motion can be referred, we have to reply that we can only do this with a limited amount of confidence. Newton himself had already pointed out that a circular motion can be detected by virtue of the centrifugal force which is thereby produced, and that no reference body is necessary [*cf.* p. 63]. But circular motion is one due to the continuous action of a force, and is therefore an accelerated motion, and in all accelerated motions the reaction due to the accelerated body is noticeable. When the motion takes place in a circle, the centrifugal force is experienced. When a railway train starts suddenly, there is a jerk which is felt by all persons in the train. This counter-effect permits us to recognize quite clearly when a body is accelerated, and apparently it is an absolute acceleration which we do discern, for we cannot expect that a stationary object would experience this counter-effect if its whole surroundings began to move or rotate.

But if the principle of relativity is to remain a general and universal law, this possibility of detecting an absolute motion must be removed. Hence in the present instance it would be necessary that the centrifugal force should arise when the water was at rest [*cf.* Newton's experiment on p. 62] and the external world around the glass was rotating. In other words, the centrifugal force would be the means of detecting not the absolute but only the relative rotation of an object. Consequently the curved path of the Trade Winds, and the rotation of the plane of vibration in Foucault's pendulum, would not only arise if, as according to **Copernicus**, the earth were rotating whilst the heavens were fixed, but also if the heavens were rotating about a fixed earth, as maintained by **Ptolemy**. It is clear that the development of this theory will include in it a new interpretation of our ideas concerning inertia. In arriving at this new hypothesis, Einstein used a fact which had long been known, but not sufficiently heeded, at least it had not been used to interpret natural phenomena; we refer to the

Reference

fact that the inertial and gravitational masses of a body are equal. By combining this fact with a conception of inertia which **Mach** had advocated fifty years before, Einstein obtained a suitable basis for his new theory, which was to surpass in its uniqueness all others which had previously been considered. We have already called attention [*cf.* p. 50] to the remarkable fact that gravity and inertia go together. Einstein made use of one of Mach's conceptions and imagined that inertia was a kind of gravitational force. This means that a force must be applied whenever an object is accelerated, because its velocity relative to other bodies in the universe is being changed. According to this theory the inertia of a body is no longer something belonging solely to that body, but it is a relationship with respect to its surroundings, in fact it depends upon its relative velocity with regard to these surroundings. Hence, whenever this relative velocity is altered, the same force must be applied as we have hitherto considered necessary to overcome the inertia of the object. This force would be equally large whether the change in velocity was caused by imparting to a top, let us say, a rotation with respect to a system of fixed stars, or conversely, if one sought to move the whole system of the fixed stars in a circle of infinitely large radius around a stationary top. For the simplest type of uniformly accelerated motion Einstein was able to show quite easily that we are no longer able to decide which of the two participating bodies is actually moving. [These two participating bodies are the moving body and the one used as a reference body.] If a person were placed inside a moving chest and did not know whether he were upon the earth or anywhere else, that person would nevertheless detect that bodies in the chest tend to move towards the floor. But he could not absolutely decide whether this phenomenon was caused by a gravitational force which attracted all objects downwards, or whether the chest was being uniformly accelerated upwards, for each of these would produce the same effect on objects inside the chest. According to ideas which had hitherto prevailed, these phenomena would be due to the reaction to a uniformly accelerated motion upwards, or to the attraction of a neighbouring planet. But because the inertial and gravitational masses are equal for any object, we cannot decide in an individual case what the action really is. In other words, we cannot say with certainty whether an accelerated motion, which we may be able to see, is made by the object or by its surroundings.

At this stage we ought perhaps to mention that **Einstein** has not yet developed this theory completely and that **Weyl,** who is otherwise an ardent exponent of these new ideas, has declared that such a completion is impossible. Moreover there are others, who, whilst they adopt the fundamental postulates of the general theory of relativity, have proposed a similar hypothesis, in which, however, it is not the surrounding objects which determine the inertia of a material body, but the æther itself. According to them a force is necessary to accelerate an object because its velocity with respect to the æther is thereby changed.

In order that this impossibility of recognizing absolute accelerations may relate to all natural processes, including all motions, and light and gravitational phenomena, the equations which represent these processes will have to be changed in many ways, since there will be things in Nature which are related to one another in a manner which is at present unsuspected. Let us consider one particular example. A ray of light SO, Fig. 198, enters a chest $ABCD$ which is moving upwards with uniform acceleration. Let the beam of light be moving in a direction perpendicular to that of *ABCD*. Whilst the light moves from *O* to *E*, let the floor of the chest move from *DE* to D_1E_1; to an observer in the chest the light will impinge upon the wall *BC* at E_1 instead of *E*. The light will have travelled along a parabolic path from which the observer can deduce his acceleration, *i.e.* the acceleration of the chest. If all this is not possible, but if, in spite of this, we are permitted to assume that the chest is standing still near to some attracting mass such as the earth, the path of the light would still be curved. It therefore follows, if the whole theory is to be retained, that a light ray likewise becomes curved when it passes near to a heavy body. If that is not so, the theory of relativity in its general form can no longer be regarded as a consistent theory. A total eclipse of the sun gives us an opportunity of observing stars whose light has skirted the edge of the sun before reaching us. If this light is deflected by the sun, then the stars will appear to be in different positions during the period of a solar eclipse than at other times. It appears, from photographs taken during recent total solar eclipses, that these predictions have been fulfilled ; even if the displacements were not numerically exact, they are at least of the correct order of magnitude.

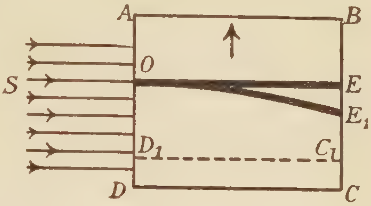


FIG. 198.

Absolute and Relative. Instead of discussing further details in Einstein's theory, or the confirmations which it has received, it is perhaps more important that the following remarks should be considered. **Newton** and his school, and **Einstein** with his followers, use the words 'absolute' and 'relative' with essentially different meanings. According to each of them, a relative motion of a body is referred to another body which contains the origin of a set of coordinates. But the way in which the surroundings influence the motion according to each school is different, in fact essentially different. Newton, whose ideas in this connexion did not differ from those of his contemporaries, maintained that the surroundings do not affect the motion of a body even when it is moving relative to them. On the contrary they are only the system of coordinates to which the motion is referred. Every motion is at least an absolute motion when considered from one point of view, but this same absolute motion is relative when referred to all other bodies having a different state of motion. It is not necessary for these other bodies to be moving ; it is only because the first is moving that a second body has a motion relative to it. It can therefore pass from a state of rest into a state of relative motion, without the least thing in it being changed. The relationship between two bodies moving relative to one another is purely a mathematical and phoronomical one. It is not necessary that there should be any kind of **physical** relationship between them. Let us assume that there is no physical connexion between the earth and the smallest and most distant of the fixed stars : *i.e.* no gravitational action, no communication by light or wireless waves. Then the fact that our earth is rotating means that it possesses a relative rotatory movement with respect to this remote fixed star.

Einstein and his adherents attach quite a different meaning to the word 'relative'. According to them, whenever a body moves relatively to another, **there is always a real influence of the two bodies upon each other** : notice the last three words. This implies that the motion is always the same from whichever of the two sides it may be viewed ; it is never absolute from one, and relative from another. No, it is always absolute from each side, absolute in the sense that it really exists. The motion of two bodies with respect to one another is similar to the mutual attraction between two electrically charged bodies, the force of attraction on one being equal to that on the other. It is for this reason that a force must be applied to set a body in motion, *i.e.* a force is necessary since the relative velocity of one body with respect to the other is being changed. Moreover, the same force must be applied whether a top is rotated with reference to the universe, or whether the universe rotates with respect to the top.

The difference between these two hypotheses may perhaps be made a little clearer if we consider the following imaginary experiment. We will suppose that in the course of some particular night the entire mass of the whole starry firmament, with the exception of one small portion, was annihilated, or that all the stars, with the exception of three small ones, were so far removed from the earth that they could be discarded. If these three stars did not happen to lie in a straight line they would suffice to determine the motion of the earth, *i.e.* the motion could be described uniquely. In other words, they would form a system of coordinates, to which the motion of the earth could be referred. Then according to Newton the motion of the earth to-day would be quite the same as it was yesterday ; and if the motion of the earth were accelerated somewhat, the same force would be necessary to-day, as yesterday. According to Einstein this is not so. The inertia of a body is due to a mutual action between the moving object and the rest of the universe. But when all matter, except for the small amount mentioned above, has been removed, the motion of the earth is different to-day, *i.e.* after the removal, from what it was yesterday. The earth would have ceased to have inertia, except for a very small fraction of that originally present. A small child could exert a force sufficient to cause the earth to leave its orbit. Relative motion has now become something absolute, for it is the carrier of real physical relationships. Consequently, to the question, ‘ Can a single body move ? ’, we must answer—it has no meaning. It is just as meaningless as the question, ‘ Does the attraction between the sun and the earth persist, after the sun has been removed ? ’ For just as two bodies are necessary before a gravitational force can be produced, and two magnetic or electric charges are necessary for the manifestation of the corresponding type of force, so can a motion only be defined if two bodies are present ; at least, that is what this new theory postulates. There are still some who have not become reconciled to these new teachings, but this is perhaps due in part to the fact that they have been brought up with the older ideas, and have not sufficiently considered this difference between the theories.

CHAPTER VII

THE AIMS OF MODERN PHYSICS

IF, in conclusion, we let our minds consider the whole field covered by physics, we shall find that there are many things which cause us to be very satisfied with the results which have been obtained. The valiant endeavours of research workers have been richly rewarded, and our insight into the secrets of Nature has been increased both in breadth and depth, especially in the latter, for the results which have been obtained during the last few decades are such that anyone who had dared even to suggest such possibilities fifty years ago, would have been regarded as a person with a vivid and somewhat fantastical imagination. Yet it is just by penetrating further into the mysteries of Nature that a few difficulties have at last been met which would appear to cause research to come to a sudden standstill. Previously to this an endeavour had always been made to discover, according to the means then approved, facts which were irreproachable, but now new facts were to hand which could not be brought into harmony with those already established. All the united endeavours of physicists in many lands were not successful in removing these difficulties, so that we may say that a real crisis had been reached in the affairs of science. Since these attempts had all ended in failure, it became necessary to reconsider the fundamental postulates which are the embodiment of all experiments, and see whether or not they did not contain something which could be improved. When such a state of affairs arises, it is useful to dismiss the difficulties of the moment in order to obtain a broader view from afar, so that by comparing our difficulties with those that have already been surmounted, we may decide on their significance, and what must be our stand with regard to them.

Scientific knowledge is very different from the knowledge of any single event or process which comes to us through our senses. Science endeavours to obtain knowledge which is quite general, and if it is dealing with natural events, an attempt is made to discover laws which govern a whole group of phenomena. And the more numerous the individual results expressed in any one law are, the greater does the significance of that law become. Amongst all the results which have led to the formulation of such laws, there are a few which stand out prominently before all others ; they are those which have enabled different groups of phenomena, apparently having no connexion with one another, to become correlated. Walls which had hitherto divided various branches of science have been broken down, and the knowledge which has been obtained in one field has been fruitful in stimulating work in others. In this way, laws have become less numerous, but those which have remained have become more significant and comprehensive. A new chapter in the history of physics has begun ! Some of the new results have already been mentioned as opportunity arose ; we shall now survey the whole field once again.

1. Long Established Facts

1. The first real progress in science began to be made at the close of the Middle Ages, when science was in a very flourishing state ; it concerned the motion of bodies. Until that time the views postulated by the ancient Greeks concerning motion had been advocated ; they only distinguished **natural** and **violent** motions. When solid bodies fell to the ground, those were natural motions, for the bodies themselves assumed the motion. The desire to fall was always latent in a solid body, so that whenever the opportunity arose, the body simply fell. The nature of a solid body was such that it wished to be placed below all liquid and gaseous matter, so that in falling this desire was being gratified. An upward motion, however, took a solid body further away from its natural position, so that this motion only occurred when the body was moved by force. From such violent motions it was maintained that nothing could be learnt concerning the nature of the body.

According to this theory the two types of motion are produced by causes which are essentially different.—But the two phenomena no longer remained isolated from each other when they began to be investigated quantitatively.* In the first place it was recognized that all bodies fell with a constant acceleration. Further investigations showed that when bodies were projected upwards the motion was uniformly retarded, and finally it was established that the numerical magnitudes of this acceleration and retardation were equal, viz. 981 cm. per sec., per sec. The two motions are therefore very closely connected with each other, in fact, the same law is applicable to each.

The celestial bodies, however, were not subject to such laws. Since these bodies are made up from some unique substance, their movements must also be quite unique ; it was therefore held that they moved along in circles. **Newton** showed, however, when the earth's survey had been completed [*cf.* p. 52], that the moon is also falling towards the earth with an acceleration which is common to all bodies, only in estimating its magnitude, due respect must be paid to the distance between the two bodies. In this way Newton showed that the force which causes a stone to fall is also operative on the moon ; he had discovered the general theory of gravitational attraction. Since the same general laws of motion are true for all bodies, we have a general theory of mechanics.

In addition, the phenomena of light, heat, electricity and magnetism were all supposed to be essentially different. That, at least, was the position in science at the commencement of the nineteenth century.

In the course of that same century, however, much progress was made. By 1820 **Oersted** had demonstrated the effect of an electric current on a magnetic needle [*cf.* p. 256]. Until then the presence of a magnetic as well as of an electric fluid had been assumed ; since then electricity and magnetism have been conceived as one.

About the middle of last century the conviction was gradually gaining ground that heat was really a manifestation of moving particles ; such a view became exceedingly probable when the mechanical equivalent of heat had been measured, and heat produced in arbitrary quantities by friction, and the kinetic theory of gases, on the other hand, had proved so successful.

---

* Historically the subject was certainly not developed in this straightforward and simple manner.

In all the instances hitherto cited, it has been a question of the correlation of two known groups of phenomena ; we now find one which is different —it concerns light. As we have already mentioned [*cf.* p. 262], **Weber** and **Kohlrausch** discovered that an optical constant, viz. the velocity of light in vacuo, was intimately connected with electricity. From this discovery **Maxwell** concluded that light waves must be electromagnetic in character, and that an oscillatory discharge from a Leyden jar would be the source of an electromagnetic wave-disturbance in the surrounding æther. We have already learned how **Hertz** actually produced these waves and found that they possessed all those properties which are characteristic of light waves. The scientific value of this discovery cannot be estimated. From the moment when this correlation between light and electricity had been obtained, the wall which had hitherto divided these two sciences disappeared, and they became as one. Other branches of physics, each of which covered quite a considerable field, also became incorporated in one another [*cf.* pp. 327 *et seq.*]; it is now known that heat rays, ultra-violet rays, Röntgen rays, the gamma-rays from radioactive bodies, and possibly also the shortest of all known rays—the so-called cosmic rays—form one long series of waves.

In broad lines these represent the discoveries of the century which has just elapsed. We now see that the whole kingdom of natural phenomena can be divided into two large branches. First, there are the phenomena due to the **mechanical motions of particles,** and this branch includes mechanics in its older sense, the theory of heat, and of acoustics. Secondly, there is that extensive field which embraces all electromagnetic phenomena. In short, we have either **particles** or **waves**. These two fields appear to be divided by an unyielding barrier. The former comprises all phenomena associated with material things, whilst the other embodies all wave-motions, which are responsible for the transmission of an enormous amount of energy across a vast space which we know to be void of matter. Material particles have an atomic structure, whereas waves must be continuous in character. This means that the two fields are separated by a deep gulf. If we once more survey the whole of physics which existed at the commencement of the present century, we have to say that research, except in a few instances, had always adhered to a straight path, leading to the discovery of important laws, which also simplified and extended our knowledge of science.

2. Have we now come to a *cul-de-sac*, beyond which research cannot advance ? Or are we still to endeavour to pursue our inquiries along the lines which have already been laid down ? In the latter case, in view of the lines along which progress has hitherto advanced, we can form a good idea as to what is required to be done. It is a question as to whether or not these two fields can become one by removing the wall which has hitherto divided them. If we pause to consider the fundamental difference between atoms and waves, we shall probably add at once that it is quite impossible, in fact it is inconceivable, that a wave and a particle should even be considered as one. But we also have to remind ourselves of the fact that an attempt to develop physics still further is really a question of fundamental changes. Moreover, we shall have to consider the basal facts of science, if we are to appreciate the significance of the step which lies ahead. It is quite certain that the development of physics is connected with the establishment of irrefutable facts ; but having discovered these facts, and investigated them so that they can be expressed quantitatively, the task of physics is not yet complete—there still remains another

and more important work to be done, and that is to discover and develop new conceptions. Even if this question has seldom been to the front during the quiet but continual progress made during the last century, it would be unpardonable if, in time of difficulties such as we are now experiencing, this question was not thoroughly examined. We shall begin by pondering over the progress which has hitherto been made, paying especial heed to this point of view.

In the first place the laws of motion were considered, and general laws obtained which were applicable to all bodies; this was very different from the teachings of the ancient Greeks, who maintained that all motions were either natural or constrained motions. How were these modifications discovered ?—Nature presents certain phenomena to us, but the interpretation of them is left to the spirit of inquiry which is present in us. The ancients were unable to deduce any primary cause for these two motions which was sufficient to account for them. Accordingly it was essential to regard natural and constrained motions as fundamentally different from one another. Thus it was not **Nature** who had erected the dividing wall between these two types of motion, it was rather due to the limited capacity of man's intellect. A definite conception having been formed in connexion with heavy bodies, all phenomena were viewed from this standpoint. There was something inherent in bodies which determined their position in the universe. It was their nature that they should rest in this position, and that they should resist any attempt made to displace them. The essential difference between a motion towards the natural position of a body, *i.e.* the centre of the earth and one which was directed away from it, was an immediate consequence of this theory. **But this difference between the two types of movement only persisted for as long as this hypothesis concerning the nature of inert matter remained.** In order to do justice to the fact that experiment had shown that all such motions were identical in their character, an entirely new conception of the material world had to be formulated. According to the new ideas of that time bodies must be indifferent with respect to their position in space, to the direction along which they move, as well as to any other velocity they may possess. Let us endeavour, for instance, to bring our minds to the task which confronted scientific men at the time of Galileo, for even now we can learn much by so doing. Even in expressing the smallest principle concerning processes in the material world, they had to conceive of them in quite another way from that to which they had been accustomed since their youth. It can be readily appreciated that most people resisted these changes of outlook, and that it took a long time before it was clear how every phenomenon was to be interpreted in terms of the new views, and until this interpretation had become, by frequent repetition, so familiar that it immediately suggested itself to the mind in later work. But after these preliminaries had been surmounted, the new views presented no further difficulty. Thus we may repeat: In terms of the old ideas of the material world, the conception of the essential identity of the two directions of motion was and remained completely incomprehensible, but it could be grasped at once, as soon as the old framework of ideas had been banished, and the new conception of bodies had been substituted.

There still remained, however, the peculiarities which were associated with the heavenly bodies. A circular motion always requires the existence of some visible constraining force, which will cause the body to deviate from its rectilinear path: a cord, band, or even a rigid wall. But there is no such obvious constraint between the earth and the moon, so that the moon was not caused to travel in a curved orbit in the same manner as a terrestrial object. If one said that it was the nature of heavenly bodies to move in circular paths, this was tantamount to acknowledging that one had not succeeded, in terms of the hitherto accepted idea of forces, in bringing the circular movement of the heavenly bodies into line with the circular motion of terrestrial bodies. It is true that the phenomena of gravitation seemed to demand a visible tie which pulled the stone downwards ; but the motion of the falling body was regarded as emanating solely from the falling body, by virtue of a ‘ something ’ the very nature of which pulled it down to its mother earth or ‘ home ’. According to their hypothesis a falling motion was not due to a mutual action between the earth and the falling body. One of the greatest services which Newton rendered to mankind was that he introduced the idea of some invisible force in connexion with his theory of gravitation, and proceeded to demonstrate its existence. From that time the motions of the celestial bodies were studied along the same lines as those of terrestrial objects. We see that the process was quite the same as in the first example. It was not Nature who had drawn a distinction between the motions of the heavenly and terrestrial bodies, but rather a distinction which man had attributed to them because he had not been able to represent both phenomena according to the existing ideas about force. The equivalence of the two phenomena only became apparent when the presence of some invisible constraining influence had been assumed.

The phenomena due to the motions of bodies on the one hand, and those due to heat on the other, especially the sensations of heat and cold which we experience from our sense of touch, are so different from one another, that it hardly seems possible for them to be essentially the same in character. But it had been known for quite a long time that we are able with our ear to perceive as sound processes which in themselves are nothing more than processes involving motion, in fact they are due to the to-and-fro movements of a tuning fork, yet the ear cannot actually notice this to-and-fro motion. Hence it was possible that heat might also be a form of motion. After it had been shown experimentally that an unlimited supply of heat could be obtained by doing work continuously, and that the kinetic theory of gases had made it probable that the molecules of a body were in a state of continual motion, the conception of heat as a special instance of motion soon began to gain ground.

Still more easily than this did it become apparent that electricity and magnetism were essentially the same in character. In this particular instance the change was not so difficult because no clear and definite ideas concerning magnetism had yet been formed. When the change was made it was not necessary to abandon any well-known and familiar facts, since magnetism played no part in the philosophy of those days : in particular it had not been regarded as intimately connected with the nature of a body, as in the case of motion.

Until the time of Maxwell, light waves had been regarded as nothing more than elastic waves. Why ? Simply because no other type of wave was known. But when experimental physics had shown the relationship between light and electricity, Maxwell pointed out the possibility that electromagnetic waves might be similar to light waves. Hertz showed that this was really their nature.

From the course of research during the last century we see that scientific knowledge is concerned with two tasks, both intimately connected with each other. The first is to obtain facts, in particular such facts as refer to quite definite properties of the objects in question, and to measure them as accurately as possible. The second is to modify continually the hypotheses, so that they always suit the facts. Then when a new theory has been formulated, it must first of all be viewed from all directions, *i.e.* all phenomena connected with this theory must be interpreted according to its postulates, and we have to make ourselves so familiar with it by using it frequently, that, in time, we are able to use it quite easily. The new theory will then appear to be quite simple and representative of the facts concerned. But in so doing, the new theory will have been extended both in breadth and depth, and when this happens real scientific progress is being made.

2. NEW FACTS

If we turn from these considerations to our former question concerning further developments in physics we shall not give so much weight to the answer which has already been given, viz. that we cannot really form any very definite views on the matter. For since present conceptions are necessarily framed in terms of theories which have hitherto prevailed, this answer cannot imply anything more than that there is no place for such further development in terms of present-day hypotheses. But that is not really the meaning of the question ; its true meaning, on the contrary, is, will it be possible to obtain another hypothesis which will enable further developments to be made ? It will therefore be quite expedient to set aside for the time being the question of the conceivability, for it concerns the second of the above-mentioned problems. The first question, on the contrary, must be, **are there facts in Nature which indicate that there may be a union or relationship between the world of atoms and the world of waves**, or are there constants which are valid for both fields, and so on ? The meaning of this statement, however, must not be falsely construed ;—science has not actually sought such facts, for they did not lie within the range of facts which all previous knowledge would have led us to expect. And when they revealed themselves of their own account, we did not at first understand them, but treated them as unwanted intruders only capable of disturbing the peace and harmony in the realm of physics, in fact, of bringing about a veritable crisis. But facts cannot be banished ; when we seek to drive them away from one place, they immediately turn up at another. However much we may endeavour to take no account of them, we are not successful.

1. The first of such facts to be met arose from investigations concerning the radiation from a black body. The laws which apply to such phenomena were not discovered until **Planck** introduced the idea of a quantum of energy. Such an idea had never before been postulated, but it became a necessity in order to discover the above laws. Soon after its introduction it was found to be indispensable in other branches of physics. According to this new theory heat rays, like matter itself, were atomic in character.

2. By making use of the idea of a quantum of energy **Bohr** was enabled to account for the emission of light rays of definite wave-lengths from various substances [*cf.* pp. 343 *et seq.*]. The quantum of energy was characteristic of each wave and only depended upon its frequency ($E = h\nu$), and was therefore a constant for the radiation emitted. But the same quantum proved itself to be a constant for the atom concerned, not only in the sense

that an atom must lose an amount of energy equal to that of the quantum which had been emitted—that would only have been a consequence of the principle of the conservation of energy.—But before the atom is capable of emitting energy, this has to be given to the atom from outside ; but the energy which an atom requires can only be received in quanta, equal to those which it can radiate out. **The quantum of energy is just as much a constant for the atom as it is for the waves.**

But a whole series of facts was known which indicated that light was a wave motion : the phenomena of interference, diffraction, and polarization are irrefutable evidence for this assumption. Moreover, light waves propagate themselves from a source through space according to the principle of Huygens. But material particles have a definite size and move in a definite direction, from which they can only be deflected by means of some external force, and when they are so deflected, it is the principle of the conservation of momentum which is valid. The new facts were strange in that some of them were such that they could not be explained in terms of the wave theory of light, but indicated that light waves had the properties of material particles—the light waves possessed momentum. On the other hand there were effects discovered in connexion with material particles which could only occur in a wave motion. The most important of these facts will now be discussed.

3. **The Photoelectric Effect.** We have already mentioned [*cf.* p. 339], that the electrons which are driven out from a metal when light of suitable wave-length is incident upon it contain the whole energy of the exciting light quantum. But part of it is spent in setting the electron free from the constraint of a metallic atom, whilst the remainder is utilized in accelerating it according to equation (167). Light from a source, however, is propagated in the form of spherical waves and fills the whole of space, whereas an atom is confined to a very small region when it receives energy. It might therefore be necessary that a metal should receive radiation for some time before a particular atom is ready to send out these liberated electrons. Actually, however, the emission of electrons occurs as soon as the light falls upon the metal. As **Einstein** first emphasized, this phenomenon is not in harmony with the idea that light is a wave motion, but rather suggests that light consists of particles which are shot into the metal ; this idea is very similar to that contained in Newton's emission theory. He called this ' needle-radiation ' ; **Silberstein** has called the entities ' light-darts '.

4. **The Compton Effect.** Another fact, which shows even more directly that light has an atomic structure, was discovered in 1923 by **Compton** ; it is known as the Compton effect. On account of the positiveness of its statements and the many definite proofs in its favour, scientists have paid considerable attention to this phenomenon even from the time when it was first discovered. The effect is a mutual action between an X-ray and an electron. When such an X-ray with its large quantum of energy strikes an electron, the action is so intense that the smaller force between the electron and the nucleus to which it is attached, can be neglected. It is as if the ray were passing through a gas consisting of electrons only [at least, this is true for atoms of low atomic weight which have a somewhat small nucleus]. What will happen when an electron is struck by a **light ray** ? The light will be diffracted, the distance of the first dark line being given by equation (194) [*cf.* Fig. 188, p. 419], viz. $ED = d = \frac{hl}{2a}$. This distance becomes greater as the diameter of the diffracting body is

reduced. Under the present circumstances when the diffracting centre is an electron, the above distance will be considerable. Fig. 199 shows the wave motion spreading outwards from the electron in a fan-like fashion.

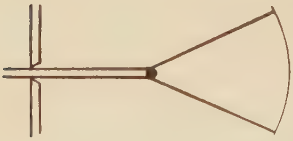


Fig. 199.

Conditions would be quite different if a particle having inertia had impinged upon the electron. Under those conditions the laws of impact would have governed the phenomena.

The electron is set in motion with a definite velocity in a definite direction. Its energy has been obtained from the colliding particle. On account of the collision having occurred, the particle which collides will, in general, be deflected from its straight line path. If, for example, the electron is only just grazed during the collision, Fig. 200, its motion afterwards will be along *b*, which is in a direction almost perpendicular to that of the colliding particle. This itself moves on with practically undiminished energy and is only deviated slightly along the direction *c*. If the collision is almost head-on, Fig. 201, the electron will experience a greater force, and move with considerable velocity in a direction which is approximately the same as that of the incident particle. This particle, itself, will also be subject to a stronger blow in consequence of which it will be driven backwards, but most of its energy will have been imparted to the electron.

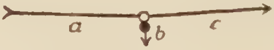


Fig. 200.

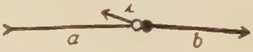


Fig. 201.

Between these two extreme cases there are many in which the impact is oblique, Fig. 202, so that after the collision the electron moves along *b*, whilst the other particle moves along *c*.

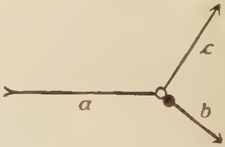


Fig. 202.

The difference between the behaviour of a wave and an inert particle in colliding with an electron is very strongly marked. When it is a wave that meets the electron a portion of a spherical wave, somewhat in the shape of a fan, will spread out from the electron, but when a material particle strikes an electron two particular directions are to be observed : they are the directions along which the particle and electron travel after the encounter. The two moving particles have the same total energy after the impact as the colliding object possessed initially.

The most remarkable result of Compton's experiment was that when he caused an X-ray, and therefore a wave, to strike an electron, the effect was just as if a **material particle had been allowed to act upon it.** There was no trace of any diffracted wave, but two directions were very clearly distinguished ; along one moved the electron which had been struck, whilst the X-ray travelled along the other. The energy of the X-ray was always found to be diminished by an amount which increased with the speed of the electron. It will perhaps be asked how such small changes in energy could be measured with such precision. Now according to the quantum theory, a quantum of energy is determined by the frequency according to the equation $e = h\nu$. But the frequency of a wave motion, or the wave-length of the radiation, can be measured very accurately ;

moreover, any change in the value of a quantum shows itself as a change in wave-length, so that we see that this change can be measured accurately. Even in his first experiments Compton was able to show that the wave-length of the X-ray after collision was less than before. The velocity of the electrons was measured by means of an electrometer,—the number of ions produced in air by the electron was determined. If ν is the frequency of the incident radiation, ν' that of the radiation after a collision had occurred, then

$$h\nu = h\nu' + \tfrac{1}{2} mv^2 \quad . \quad . \quad . \quad . \quad (203)$$

where m and v are the mass and velocity of the electron respectively.

These experiments have been repeated and the results confirmed by several other workers. In some recent experiments by **Bothe** and **Geiger** two detecting devices, similar to that used in the work described on page 318, have been employed. These devices consisted of a small metallic box inside of which there is an insulated needle. One of these ionization chambers was arranged to receive the electron, whilst the X-ray entered the other. The deflections of the electrometer were recorded automatically, so that the moment when an electron or X-ray entered the chambers could be ascertained. It was found that whenever an electron entered the one box, an X-ray had passed into the other. We therefore have to conclude that although numerous phenomena indicate that X-rays are a wave motion, yet when they collide with electrons they appear to possess **inertia** and **momentum**, *i.e.* they exhibit the properties of a material particle.

5. The above phenomena would become substantially more important and the deductions more convincing, if the converse phenomena could be discovered, *i.e.* to show that material particles had properties peculiar to a wave motion. It has already been shown [*cf.* p. 415], that diffraction effects are due to a wave motion. When light falls on a grating diffraction effects are produced. If the diffracting centre consists of a number of parallel and equidistant lines ruled on a surface, the diffraction pattern consists of a number of lines, each line being a diffracted image of the slit. But if the diffracting centre is a point, then the pattern produced consists of a number of concentric rings. In order to measure the wave-length of X-rays it will be recalled that a crystal was used as a grating.

In 1927 **Davisson** and **Germer** caused not waves, but electrons, and therefore small material particles, to fall upon a nickel crystal. They found that, after reflexion, phenomena previously attributed to waves could be detected. **G. P. Thomson** has repeated these experiments, but he allowed the electrons to pass through thin sheets of gold or aluminium foil. Diffraction rings similar to those produced by X-rays were observed. In 1928 **Rupp** was able to confirm these experiments.

3. MODERN AIMS

When we compare previous hypotheses with their sharp distinction between waves and particles, with the facts, we can scarcely maintain that they always do justice to the facts. According to these theories it would be impossible for waves and particles to have properties common to each other. This was the reason why, in these days, scientists have argued that a new theory must be sought, in which all these phenomena will be accounted for correctly, *i.e.* the gulf between waves and quantum must be bridged.

The means whereby this might be accomplished have been two: the first attempt is referred to as the **wave-mechanics.** In it the wave is regarded as the primary element and an attempt is made to account for the quantum theory in terms of these waves. The second attempt is known as the **quantum-mechanics.** It assumes the smallest indivisible entity—the quantum—to be the fundamental element and has shown that all wave phenomena are properties of the quantum.

1. The wave-mechanics was initiated by **de Broglie,** but was considerably extended and developed by **Schrödinger.** Even in 1924 de Broglie had assumed that not only did waves exhibit quantum properties, but that material particles had a twofold nature, and that they were always associated with waves. Since an energy E can be expressed according to equation (153) as $E = mc^2$, it might also be equal to $h\nu$, where h is Planck's constant. He therefore wrote,

$$h\nu = mc^2 \quad . \quad . \quad . \quad . \quad . \quad (204)$$

and from the known mass m of the electrons and the nucleus he calculated the frequency ν for a vibration which is inseparably associated with this energy quantum. In the above equation c is the velocity of light. For an electron of mass $m = 9 \times 10^{-28}$ grm., and $c = 3 \times 10^{10}$, we have $\nu = 1{\cdot}24 \times 10^{20}$ per second. For the nucleus of a hydrogen atom we have $\nu = 2{\cdot}3 \times 10^{23}$ per second. It therefore follows that a wave train is always attached to a moving body, just like the water waves around a moving ship, or the air waves which accompany a projectile moving through the atmosphere, if the analogy is not carried too far. Amongst all the waves which are associated with a moving body, there will be points at which the planes of many of the individual waves are equal. Energy will tend to become concentrated at these points and this energy is equal to that of the moving particle, whilst the velocity with which such a point moves forward will equal that of the moving particle. De Broglie was able to show that these energy-nodes will exist for some time: it was this fact which enabled such a theory to be adopted. In a similar way all those properties, which had hitherto been regarded as mechanical properties of a body, must now be considered as characteristics of this 'wave-packet'. Above all it can be shown that when such a 'wave-packet' moves in a circle, it can only do so in orbits the length of which is an integral multiple of that of the wave-length concerned, because in all other orbits interference causes the waves to be destroyed. The orbits which Bohr had assumed in his theory were now a necessary condition of the new theory. [More recently some doubt has been expressed concerning the validity of these conclusions.]

Schrödinger has extended these ideas and so developed a more complete wave-mechanics. His chief service in this respect was to construct those mathematical tools with the help of which a whole series of facts have been explained according to this theory. Even the diffraction phenomena observed by Davisson and Germer are assumed to be caused by these material waves. It is thought that they are due to the waves postulated by de Broglie, especially since their observed wave-lengths agree with those calculated by theory.

2. Shortly before Schrödinger had published his new theory, **Heisenberg** had been developing the quantum-mechanics. His starting point was the quantum. In working out this theory, he was guided by the principle that only those quantities should be employed in it which could be detected and measured accurately. First and foremost these are the energy

quanta. They have been shown to be independent of all theories—we refer especially to the work of Franck and Hertz in this connexion. Moreover, the wave-lengths of the various electromagnetic waves have all been measured accurately. They are related to the energy quanta by the following relationship due to Bohr, viz.,

$$E_1 - E_2 = h\nu \quad . \quad . \quad . \quad . \quad . \quad (205)$$

This equation was the foundation from which Heisenberg was able to obtain the wave-lengths from the energy quanta. Finally it is possible to measure the relative intensities of the various lines in a spectrum. Heisenberg interpreted these as a measure of a probability, and therefore of the frequency with which an electron made a jump from one orbit to another.

But what has not been proved, and what cannot be proved as knowledge now stands, is the existence of the electronic orbits around the nucleus, as postulated by Bohr. Moreover, we can never learn at what definite time an electron will be at a certain point in an orbit. We only know that there are different states to an atom, *i.e.* the energy contents of an atom vary from time to time : the variations are always equal to a quantum. That these differences are due to the passage of an electron from one orbit to another is only an assumption due to Bohr. Hence these quantum orbits play no part in Heisenberg's theory : they only remain as energy levels. In general, Heisenberg has shown that the happenings inside an atom cannot be stated explicitly. One electron cannot be differentiated from another, so that they cannot be traced in one orbit and recognized when they appear in another. We do not know what is the shape of an orbit, nor where the electron is situated in it.

It is sometimes stated, in view of the above remarks, that the behaviour inside an atom is not causal ; such a statement is at least very misleading. This law of causality says, in its application to physical processes, that we can specify the condition of a material object at some future time, if we know it at some earlier time. In the above instance we never know the condition at any one time, so that we cannot express it at any other time. This only means that we cannot apply this principle in these circumstances. Even if we were sure that we could never be able to express these states or conditions exactly, yet there is a very great difference between whether the causality of a process can be proved, or whether a process happens according to the law of causality.

With this decision not to use the idea of electron orbits, the whole picturesqueness of Bohr's theory has to be given up. Even though we can agree with Heisenberg, that the progress of science should not be made to depend upon mental pictures, it is nevertheless in part a matter of chance, what we have learned to measure exactly, and what we have not so learned. Moreover, it is still possible that the problem would be rendered unnecessarily difficult if the description had to contain only those magnitudes which have hitherto been measured and observed. But in his recent publications, Heisenberg has departed somewhat from this rigid attitude.

Heisenberg, also, has already succeeded in interpreting many phenomena according to the tenets of his theory. Many other scientists are now working along similar lines—**Born** and **Jordan** in Germany, and **Dirac** in England. In spite of the great difference between the starting points in the wave-mechanics, and in the quantum-mechanics, the results obtained by the two theories are consistent with one another. But at the present time neither of these two theories has been so far developed, that it can be said that the goal has been reached. At the moment everything is, so to say, in a state of flux. Many new ideas are still embodied in the theory, but also the meanings of the new views are frequently being modified. Under these circumstances any attempt to estimate the value of the new theories must be considered premature.

It is even conceivable that there are many, who, in their desire to see the goal reached, will regard the progress as far too slow. But one must remember how very great and difficult are the tasks which confront physical science to-day. The great dividing wall between the world of atoms and the world of waves has to be broken down. The whole structure of physics is involved. Any further development of natural laws appears to be impossible. The first part of the work has been done,—atomic characteristics have been shown to exist in waves, and waves in particles. The question now confronting us is of a twofold nature : Why is it that a wave can have atomic properties, and a particle the characteristics of a wave ? What is the significance of the incidence of each of these into the fundamental phenomena of the other ? The answer to this question will have to explain at once the secrets of light as well as of the material world ; it will encompass the happenings of the whole infinite universe. Such is the magnitude of the task, such is its difficulty.

INDEX TO SUBJECTS

Aberration, chromatic, 402 ; — of light, 385 ; — spherical, 397
Absolute and relative, 454
Absolute zero, 147
Absorption, 334 ; — spectrum, 423
Acceleration, 21
Action and reaction, 34 ; — mutual, between currents, 262
Adhesive forces, 131
Æther, 23, 60, 382, 389, 437 ; — motion of, 441
Aims of modern physics, 453
Alpha particles, 315 ; — counting of, 317 ; — tracks of, 319
Atomic constants, 204 ; — heats, 157 ; — at low temperatures, 158 ; — numbers, 369 ; — theory, confirmations of, 317
Atoms, 201, 380 ; — mass of, 212 ; — number of, 204 ; — structure of, 221, 367
Attraction, electric, 240 ; — law of, 53 ; — nature of, 59
Avogadro's hypothesis, 126

Balmer's formula, 342
Biot's law, 258
Biprism, 413
Black-body radiation, 335
Bohr's theory, 346
Boiling, 171
Boyle's law, 121
Brownian movement, 133, 209

Calorimeter, animal, 183 ; — bomb, 182
Cathode rays, 287
Cells, galvanic, 284
Chromatic scale, 90
Circuits, closed, 250
Cohesive forces, 130
Colour, 434
Compton effect, 459
Condensers, 240
Conduction, metallic, theory of, 251
Conductivity equivalent, 281
Coolidge tubes, 297, 358
Coordinates, 12
Correspondence principle, 346
Coulomb's law, 234
Crystals, 136 ; —, systems of, 137
Current, electric, 246 ; — saturation, 303

Diamagnetism, 232, 375
Diameters of molecules, 211
Diatonic scale, 89
Dielectric constant, 242
Diffraction, 415 ; — at a circular aperture, 417 ; — gratings, 419
Diffusion, 132
Dimensions, theory of, 19
Dispersion, 399 ; — anomalous, 404 ; — theory, 403
Dissociation, electrolytic, 279
Doppler effect, 91

ε, 292
Echo, 96
Elasticity, bulk, 129 ; — Hooke's law of, 66 ; —, due to bending, 67 ; — longitudinal, 65 ; — modulus of, 66 ; —, torsional, 67
Electricity, distribution of, 236
Electrolysis, 270
Electromagnetics, 256
Electrons, 287, 294 ; — mass of, 297
Electroscope, 233
Electrostatics, 232
Elements, classification of, 109
Emission, 334 ; — of radiation from excited atoms, 352 ; — of Röntgen rays, 364 ; — spectrum, 423
Energy, 37 ; — conservation of, 176 ; —, kinetic, 40 ; —, potential, 39
Entropy, 185 ; — function, 189 ; — law as a statistical law, 195 ; — and probability, 195
Evaporation, 66
Exclusion principle, 376
Expansion of gases, 144 ; — isothermal, 191 ; — liquids, 144 ; — solids, 143

Films, interference in, 408 ; — monomolecular, 202
Force and motion, 25 ; — centrifugal, 43
Frequency, 76
Fresnel's mirrors, 413
Fusion, 164 ; — heats of, 165

Galvanometer, 261 ; — string, 261 ; — tangent, 261
Gas equation, 147
Gases, 113 ; — diffusion of, 123 ; — kinetic theory of, 116 ; — pressure of, 115
Gram-equivalent, 281 ; — molecule, 208
Gravitation, law of, 53
Gravitational constant, 54 ; — fields, 50
Gravity, 48, 438

Hot-wire instruments, 249

Ice pail, 239
Impact, 69 ; —, elastic, 70 ; —, inelastic, 70 ; —, oblique, 76
Impulse, 33
Induction, due to currents, 264 ; — electromagnetic, explanation of, 268
Inertia, and gravitation, 62
Influence, electric, 238
Interference, 408 ; — of polarized light, 427
Interval, musical, 89
Ionic collisions, 305 ; —, mobilities, 281
Ionization potentials, 305
Ions, gaseous, 301
Isotopes, 215, 217 ; — atomic weights of, 218

Joule's law, 249

Law of energy, universality of, 181
Laws of electrolysis, 271
Length, measurement of, 7; — standard, 8; — absolute, 9
Lenses, 400
Light, natural, 432; — propagation of, 438; — relative, 9; theories of, 387; — velocity of, 385
Lines of force, electrostatic, 244; — magnetic, 227
Liquids, pressure of, 129; — shape of, 128; — true volume of, 129

Magnetic fluids, 229; — potential, 224; — units, 224
Magnetism, 222; — theories of, 229
Magneton, 374
Mass and motion, 28; — spectrograph, 216
Mean free-path, 205
Mechanical forces in electrolysis, 283; — equivalent of heat, 180
Medium, influence of, on electricity, 241; — on gravitation, 60; — magnets, 225
Microscope, 406
Molecular constants of gases, 207; — magnets, 230
Molecule, diameter of, 211; — kinetic energy of, 212
Momentum and impulse, 33
Motion, 12; —, absolute and relative, 15; — harmonic, 44; — in a circle, 41; —, laws of, 24; —, simple and compound, 16

Newton's rings, 408
Nodes, 78

Ohm's law, 247
Orbits, electron, 353; — elliptical, 356
Oscillation, period of, 46; — damped, 47
Oscillations, electrical, 267
Osmosis, 132; osmotic pressure, 132

Paramagnetism, 232, 373
Passage of rays through solids, 306
Peltier effect, 255
Perpetual motion, 189
Phase, 46; — effect of, 395
Photoelectric effect, 295, 459
Pitch, 99
Polarization, 75; — electric, 243; — of light, 425
Poles of a magnet, 223; — strength of, 223
Positive rays, 299
Potential, 56; — electric, 235; — of the earth, 237
Pressure, critical, 173
Prism, refraction by, 399
Probability, 119
Processes, positive and negative, 187
Proton, 379

Quantum mechanics, 462
Quantum theory, 336; —, and atomic heats, 159; —, confirmation of, 339; —, and periodic system, 369

Radioactive changes, 377; — decay, 313; — substances, 313
Radioactivity, 308
Radium, 309
Rays, 87; — from radioactive bodies, 309; —, heat, 331; —, light, 331
Reflexion of light, 392; — spherical waves, 85
Refraction, 85, 399; — double, 427
Relativity, Einstein's, 447; — general theory of, 449; Newton's, 31
Resistance, 248
Resistivity, 248
Resolving power, 420
Resonance board, 100
Rocker, Trevelyan, 91
Röntgen rays, 139
Rutherford's atom, 324

Screening, electric, 239
Secondary processes in electrolysis, 274
Self-induction, 267
Siren, 90
Sound, intensity of, 93; —, velocity of, 93; —, —, effect of wind on, 95; — — in solids, 98
Sources of light, weak, 407
Space, closed, 29
Specific heat, 150; — of gases, 152
Spectra of gases, 340
Spectrum analysis, 422; — characteristic, 36; continuous, 360
States of aggregation, 162
Strings, vibrating, 101
Sublimation, 170
Surface, equipotential, 58
System, periodic, 110
Systems of electrical units, 262

Temperature, 141; — critical, 173; — and radiation, 332
Theories of electrolysis, 278
Thermal capacity, 151
Thermions, 296
Thermodynamics, laws of, 174; — third law of, 197
Thermoelectricity, 254
Thermometers, 142
Thomson effect, in gases, 125
Timbre, 93, 103
Time, measurement of, 9; — unit, 11
Transport members, 275

Ultimate thermal end of universe, 196
Units, absolute, 19; — derived, 18; — practical system of, 263

Vapour pressures, 168
Vaporization, heats of, 171
Velocity, 17; —, molecular, in gases, 202, 213; — relative, 441; — unit of, 17; —, variable, 20
Vibration, period of, 433
Vision, defects of, 403
Voltameters, 273

Wave, motion, compound, 77; — mechanics, 462; — packet, 462
Waves, 73; — electromagnetic, 327, 386; — longitudinal and transverse, 75; —, stationary, 77, 414
Wind instruments, 101
Work, 37

X-rays, 358; — secondary, 365

Zones, Fresnel, 82

INDEX TO AUTHORS

Abe, 403
Abraham, 298
Adam, 202
Alhazen, 387
Ampère, 256
Arago, 430
Aristotle, 23, 106, 387, 437
Arrhenius, 279
Aston, 216, 300
Atwater, 183
Atwood, 21
Avogadro, 126

Balmer, 342
Barkla, 361
Bartholinus, 427
Becquerel, 308
Bernoulli, 116
Berzelius, 107
Bessel, 50
Beudant, 97
Biot, 258
Black, 151
Bohr, 342, 369, 458
Boltzmann, 116, 119, 154, 161, 192, 335
Borelli, 51
Born, 462
Bosković, 9
Bothe, 461
Bowen, 373
Boyle, 68, 114, 408
Boys, 55
Bradley, 385
Braun, C., 233
Braun, F., 55
Brewster, 426
de Broglie, 462
Brown, 133
Buchner, 297
Bunsen, 341, 423
Buridanus, 49

Cagniard la Tour, 89
Carnot, 189
Cauchy, 405
Cavendish, 54
Celsius, 143
Christiansen, 404
Clapeyron, 147
Clausius, 116, 190, 251
Colladon, 98
Compton, 375, 459
Coolidge, 297, 358
Copernicus, 51, 449
Coulomb, 223, 234, 301
Crookes, 391
Curie, M., 308, 381
Curie, P., 308

Dalton, 201, 215
Daniell, 285
Da Vinci, 49
Davisson, 461
Debye, 158
De Haas, 258
Democritus, 23, 105
Descartes, 385
Devaux, 202
De Vries, 133
Dirac, 462
Doppler, 91
Drude, 252
Duhem, 48
Dulong, 157
Dunoyer, 203

Einstein, 50, 63, 209, 258, 298, 339, 446, 450
Elster and Geitel, 295
Empedocles, 106
Eötvös, 50
Eucken, 159
Euler, 434
Ewing, 231
Exner, 233

Fahrenheit, 142
Fajans, 378
Faraday, 202, 226, 239, 241, 271, 438
Feddersen, 329
Fermi, 376
Fizeau, 386, 412, 446
Fleming, 260
Foucault, 386, 388
Fourier, 77
Fowler, 349, 373
Franck, 349, 355
Franz, 251
Fraunhofer, 423
Fresnel, 81, 384, 413, 429
Friedrich, 359

Galileo, 24, 48
Gans, 374
Gassendi, 116
Gauss, 213, 224
Gay-Lussac, 147
Gehrke, 412
Geiger, 204, 318, 324
Geissler, 306
Geitel, 295
Gerber, 61
Gerlach, 375
Germer, 461
Giese, 301
Gilbert, 222
Goldstein, 299

Gomperz, 24
Goudsmit, 375
Griesinger, 179
Grimaldi, 415
Grotthus, v, 278

Halley, 51
Harkins, 380
Heisenberg, 462
Helmholtz, 175, 198, 273
Hepperger, 61
Hertz, G., 349
Hertz, H., 43, 328, 386, 455
Hess, 319
Heusler, 230
Hittorf, 275
Honda, 374
Hönigschmid, 216
Hooke, 51, 66, 408
Huygens, 41, 51, 81, 384, 387, 416, 429

Isenkrahe, 61

Jordan, 462
Joule, 125, 175, 180, 249

Kamerlingh-Onnes, 197, 253, 381
Kant, 10
Kaufmann, 297
Kelvin, *vide* Thomson
Kepler, 51
Ketteler, 405
Kirchhoff, 334, 341, 423
Kirsch, 324
Knipping, 301, 359
Kohlrausch, F., 80
Kohlrausch, R., 262, 280
König, 100
Kossel, 370, 381
Kundt, 155

Langevin, 373
Langmuir, 202, 381
Laplace, 61
Laue, 139, 301, 359
Lavoisier, 106
Lawson, 319
Lenard, 306
Leucippus, 105
Lewis, 370, 381
Linde, 125
Lorentz, H. A., 268, 298, 330
Loschmidt, 208, 339
Lummer, 412

Mach, 62
Malus, 425, 429
Mariotte, 114
Marsden, 324
Marx, 359, 386
Maxwell, 80, 116, 119, 123, 207, 213, 245, 262, 328
Mayer, J. R., 175
Meitner, 324
Mendelejeff, 109
Mersenne, 88, 94
Meyer, L., 109
Michelson, 442, 446
Mie, 352
Millikan, 292, 300, 339, 373
Minkowski, 14
Moseley, 361

Nernst, 158, 199, 284
Neumann, 298
Newton, 24, 31, 34, 41, 49, 51, 59, 62, 387, 400, 408, 436, 450
Nollet, 132

Occam, 48
Oersted, 256
Ohm, 247
Oppenheim, 61
Oresme, 49
Ostwald, 201

Pauli, 376
Peltier, 255
Perrin, 209, 301
Petit, 157
Pettersson, 324
Pfeffer, 132
Philoponus, 48, 51
Pickering, 349
Planck, 336, 458
Plato, 23, 106
Pliny, 222
Proust, 107
Prout, 107, 391
Ptolemy, 449
Pythagoras, 88

Ramsay, 110, 315, 424
Regnault, 123, 150
Riecke, 251, 268
Ritz, 356
Robison, 89
Römer, 385
Röntgen, 301
Rowland, 257, 419
Royds, 315
Rubner, 183
Rupp, 461
Rutherford, 163, 204, 309, 312, 317, 323, 343, 381

Savart, 90, 258
Schäfer, 298
Schrödinger, 462
Snell, 86
Soddy, 312, 378
Sommerfeld, 253, 357, 379
Soto, Dominicus, 49
Stark, 36, 330, 360
Stefan, 335, 338
Stern, 203
Stokes, 440
Straton, 113
Sturm, 98

Tempier, 48
Thomson, G. P., 461
Thomson, J. J., 289

Thomson, W. (Kelvin), 125, 189, 192
Tolman, 252
Trevelyan, 91

Uhlenbeck, 375

Van der Waals, 123, 149, 209
Van't Hoff, 133
da Vinci, 49

Wagner, 360
Warburg, 155, 355
Weber, 80, 158, 230, 262
Weiss, 374
Weyl, 450
Wiedemann, 251
Wien, W., 335, 360
Wiener, 414
Wilson, C. T. R., 204, 320
Wolz, 297
Wren, 51
Wulf, 234

Young, 425

Zeeman, 36, 330
Zeiss, 403

BOSTON PUBLIC LIBRARY

Printed in Great Britain by
Butler & Tanner Ltd.,
Frome and London

AR.